The Dissertation Committee for Paul John Paukstelis Certifies that this is the approved version of the following dissertation:

Structural studies of a group I intron splicing factor and a continuous three-dimensional DNA lattice.

Committee:

Alan M. Lambowitz, Supervisor

Karen S. Browning

David W. Hoffman

Jon D. Robertus

John J.G. Tesmer

Structural studies of a group I intron splicing factor and a continuous three-dimensional DNA lattice.

by

Paul John Paukstelis, B.S.

Dissertation

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

The University of Texas at Austin
May, 2005

Dedication

To my father, Joseph V. Pauk_telis, Ph.D.

Acknowledgements

I would like to thank the following people: Dr. Alan Lambowitz for the support and the opportunities he has given me, Drs. Andrew Ellington, and Edward Marcotte, for the their support and encouragement on the DNA work, Drs. Sabine Mohr and Georg Mohr for their insight and helpful ideas, Robert Coon and Tony Kaiser for their work on the CYT-18 biochemistry, Dr. Lakshmi Madabusi for starting the CYT-18 crystallization project, Dr. Jacek Nowakowski for getting me started in structural biology, Dr. John Tesmer for answering my crystallography questions, my classmates and friends, Arun Ramani, Mike Karberg, Shailesh Date, and Alex Nakorchevsky, for the Friday croquet and Crown & Anchor sessions, Kate Monzo for keeping me laughing, and my family for their support and for making me one of the luckiest people on Earth.

Structural studies of a group I intron splicing factor and a continuous three-dimensional DNA lattice

Publication No._____________

Paul John Paukstelis, Ph.D.
The University of Texas at Austin, 2005

Supervisor: Alan M. Lambowitz

The *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase (CYT-18 protein) functions in splicing group I introns by stabilizing the catalytically active structure of the intron RNA. I determined 1.95 Å X-ray crystal structure of a C-terminally truncated CYT-18 protein (ΔC423-669) that efficiently splices the *Neurospora crassa ND1* intron and other group I introns. The structure shows that CYT-18's nucleotide-binding fold and intermediate α-helical domain are essentially the same as those of closely related bacterial TyrRSs, except for an α-helical N-terminal extension (H0) and two small insertions (I and II) in the nucleotide-binding fold. The X-ray crystal structure in conjunction with site-directed hydroxyl radical cleavage data enabled the construction of a refined model of the CYT-18/group I intron RNA complex. The model shows that the group I intron RNA, like $tRNA^{Tyr}$, binds across the surface of the two subunits of the homodimer, but interacts with the side opposite the aminoacylation active site. Though CYT-18 contains a $tRNA^{Tyr}$ binding site, and there are structural similarities between group I introns and tRNAs, these results demonstrate that CYT-18 adapted to function in intron splicing by acquiring a distinct binding site for group I introns.

DNA has proved to be a versatile material for the rational design and assembly of nanometer scale objects. I solved the crystal structure of a continuous three-dimensional DNA lattice formed by the self-assembly of a DNA 13-mer to 2.1 Å resolution. The structure consists of stacked layers of parallel helices with adjacent layers linked through parallel-stranded base pairing. The hexagonal lattice geometry contains solvent channels large enough to allow 3'-linked guest molecules into the crystal. I have successfully used these parallel base pairs to design and produce crystals with greatly enlarged solvent channels. This lattice may have applications as a molecular scaffold for structure determination of guest molecules, as a molecular sieve, or in the assembly of molecular electronics. Predictable non-Watson-Crick base pairs may serve as a new tool in structural DNA nanotechnology.

Table of Contents

List of Tables

List of Figures

PART I: A TYROSYL-TRNA SYNTHETASE ADAPATED TO FUNCTION IN GROUP I INTRON SPLICING BY ACQUIRING A NEW RNA BINDING SURFACE

Chapter 1. Introduction

1.1 GROUP I INTRONS

Proper excision of non-coding regions (introns) from coding regions (exons) of RNAs is a critical step in gene expression (Cech & Bass, 1986; Cech, 1990). Several classes of introns using different splicing mechanisms have been identified, including group I, group II, tRNA, archaeal, and spliceosomal introns (reviewed in Cech & Bass, 1986).

Group I introns are the best understood class of introns following two decades of extensive biochemical investigation. Several recent crystal structures have confirmed and clarified many of the details of their three-dimensional structure and catalytic mechanism (Adams et al., 2004; Guo et al., 2004; Golden et al., 2005). Group I introns were the first RNAs demonstrated to be catalytically active (Cech et al., 1981; Zaug & Cech, 1982). They splice using a series of trans-esterification reactions to remove themselves from protein-coding regions of messenger RNAs, ribosomal RNAs, and tRNAs (Haugen et al., 2005). This process begins with the attack of the 3' hydroxyl group of a guanosine cofactor at the 5' splice site and is followed by the attack of the 3' hydroxyl of the 5' exon on the 3' exon. The final products are ligated exons and excised intron (Figure 1.1). Though different group I introns have limited sequence homology, they share conserved

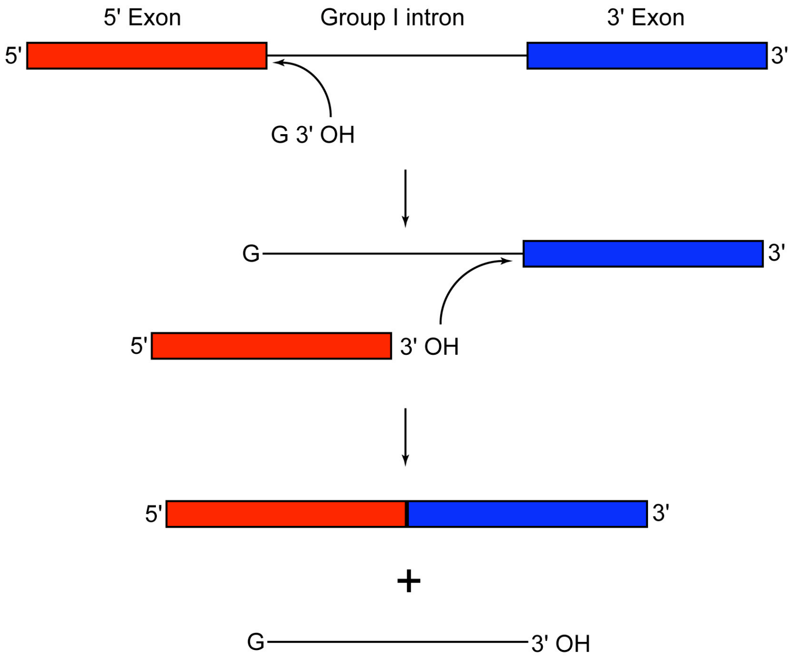

Figure 1.1. Schematic representation of group I intron splicing. Group I introns are removed from flanking RNA sequences through a series of trans-esterifcation reactions. An exogenous guanosine nucleotide initiates attack at 5' exon (red). The 3' hydoxyl group of the 5' exon is now free to attack the 3' exon (blue). The final products are ligated exons and free intron.

secondary and tertiary structures and utilize the same splicing mechanism.

The conserved secondary structure of group I introns consist of base paired regions (P), junction regions (J) between base-paired regions, and loops (L) that cap base-paired stems (Figure 1.2). The group I intron catalytic core is composed of two extended helical domains, denoted P4-P6 and P3-P9, which are juxtaposed to form the active-site cleft (Michel and Westhof, 1990; Golden et al., 1998; Adams et al., 2004; Guo et al., 2004; Golden et al., 2005; Figure 1.2). In addition to the P4-P6 and P3-P9 domains, group I introns can also contain a number of peripheral elements that facilitate the proper orientation of the core helices to form the active site (Michel and Westhof, 1990). Several key tertiary interactions between the two domains of the catalytic core and between the catalytic core and peripheral elements are important for proper folding. Specifically, tetraloop-tetraloop receptor interactions between L9 and P5 help stabilize and orient the two core helices (Michel and Westhof, 1990), while interactions between L2 and P8 and a series of interactions between P5abc and the P4-P6 domain provide additional stability to the catalytic core (Costa and Michel, 1997; Batey et al., 1999).

Many group I introns can self-splice *in vitro*, requiring only the addition of magnesium and guanosine for the reaction, while other group I introns require protein cofactors for efficient splicing. Most if not all group I introns are likely to rely on proteins for proper folding and assembly *in vivo* (Lambowitz and Perlman, 1990). Protein-assisted group I intron splicing provides a unique system for studying RNA-protein interactions, RNA catalysis, and the evolutionary transition from the RNA-world to the RNA/protein-world.

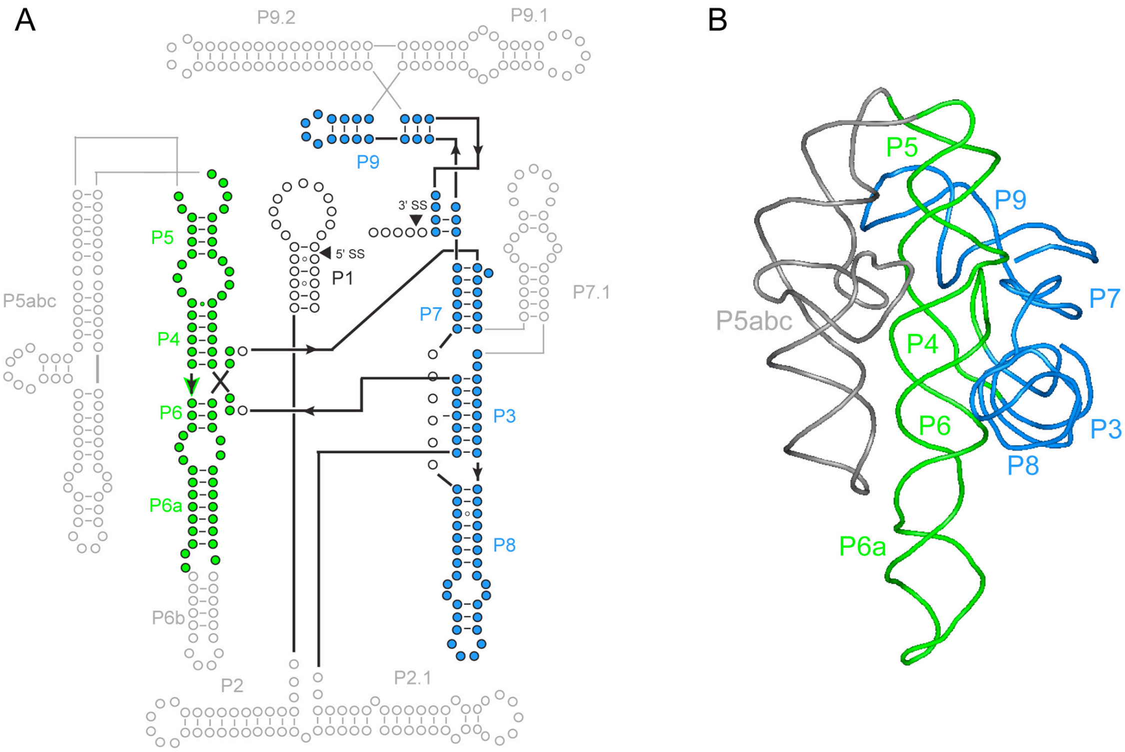

Figure 1.2. Two- and three-dimensional representations of group I introns. (A) A secondary structure representation of a group I intron showing the base pairing interactions that make up the domains of the intron. The P4-P6 domain is shown in green, the P3-P9 domain in shown in blue, and a variety of peripheral extensions are shown and labeled in gray. (B) The crystal structure of the *Tetrahymena thermophila* LSU intron showing the relative orientation of the P4-P6 and P3-P9 domains colored as in (A) (Golden et al., 1998). The P5abc peripheral element is shown in gray.

1.2 CYT-18 AND PROTEIN-DEPENDENT GROUP I INTRON SPLICING

The *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase (mt TyrRS), which is encoded by the nuclear gene *cyt-18*, is a bifunctional enzyme that catalyzes the aminoacylation of mt $\mathrm{tRNA}^{\mathrm{Tyr}}$ and promotes the splicing of mt group I introns (Akins and Lambowitz, 1987). A point mutation in the *cyt-18* gene (*cyt-18-1*; G127E) results in temperature-sensitive defects in both mt TyrRS activity and the splicing of several mt group I introns, and these defects are separable by secondary mutations that restore splicing but not TyrRS activity (Mannella et al., 1979; Akins and Lambowitz, 1987; Cherniack et al., 1990). CYT-18 has been shown to be required for the splicing of at least three group I introns in *Neurospora* mitochondria: the large ribosomal subunit rRNA intron (mt LSU), the *ND1* intron from the NADH reductase subunit I gene, and *cob*-I2 from the cytochrome *b* gene (Garriga and Lambowitz, 1986; Wallweber et al., 1997). Biochemical analysis showed that purified CYT-18 protein can promote the *in vitro* splicing of these and other diverse group I introns, that it recognizes primarily a conserved tertiary structure of group I intron RNAs, and that it functions in splicing by stabilizing the catalytically active RNA structure *(*Guo et al., 1992; Mohr et al., 1992; Caprara et al., 1996a,b).

The CYT-18 protein is a class I aminoacyl-tRNA synthetase (aaRS) homologous to well-studied bacterial TyrRSs (Figure 1.3). These proteins consist of an N-terminal nucleotide-binding fold domain, which binds the tRNA's acceptor stem and catalyzes aminoacylation, followed by intermediate α-helical and C-terminal domains, which bind the tRNA's anticodon and variable arms (Carter et. al., 1986). The C-terminal domain is

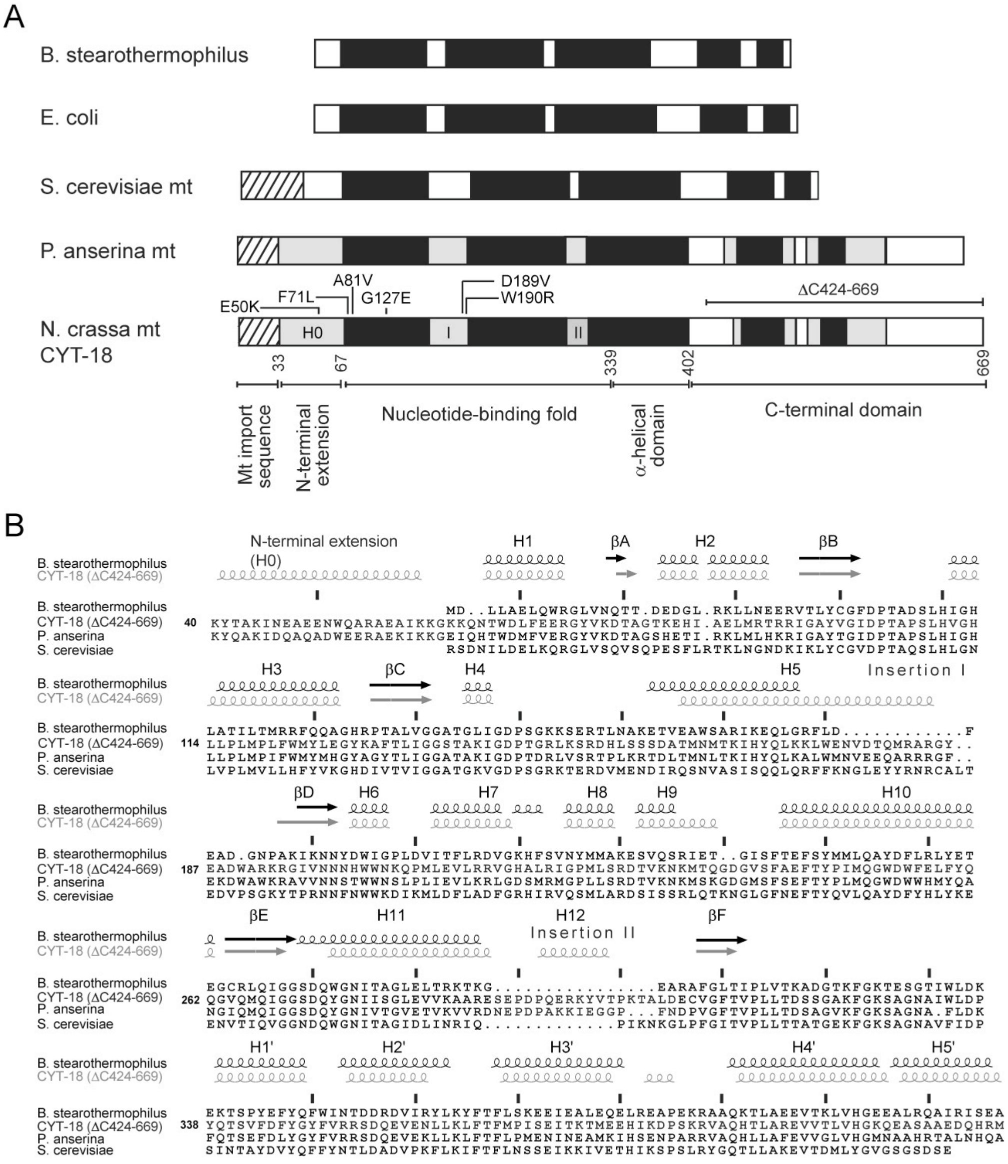

Figure 1.3 Comparison of CYT-18 with bacterial and mt TyrRSs. (A) Schematic showing protein regions. Regions conserved in all bacterial and organelle TyrRSs are in black, and insertions in splicing-competent TyrRSs are in gray. The locations of the *cyt-18-1* mutation (G127E) and second-site suppressor mutations that restore splicing but not TyrRS activity are indicated above CYT-18. (B) Structure-based amino acid sequence alignments of CYT-18 and *B. stearothermophilus* TyrRS generated by EndScript 1.0 (Gouet et al., 1999), with sequence-based alignments of the *P. anserina*, and *S. cerevisiae* mt TyrRSs shown below.

connected to the remainder of the protein via a flexible linker and contains a region structurally related to *E. coli* ribosomal protein S4 (Guijarro et al., 2002). The bacterial TyrRSs are homodimers, which bind and aminoacylate one tRNA^{Tyr} molecule ("half-of-sites reactivity"; Jakes and Fersht, 1975). The tRNA binds across the two subunits of the homodimer, with the nucleotide-binding fold of one subunit interacting with the tRNA's acceptor stem, and the intermediate α-helical and C-terminal domains of the other subunit interacting with its anticodon and variable arms (Winter et al., 1983; Carter et al., 1986; Yaremchuk et al. 2002). TyrRSs differ from other class I aaRSs in interacting with the major groove of the acceptor stem, similar to class II aaRSs (Bedouelle and Winter, 1986; Yaremchuk et al., 2002).

Like bacterial TyrRSs, CYT-18 is a homodimer, which binds one molecule of group I intron RNA (Saldanha et al., 1995). The nucleotide-binding fold and intermediate α-helical domain of CYT-18 are the minimal regions required for splicing the *N. crassa ND1* intron and most other group I introns. Although CYT-18's C-terminal domain contributes to binding group I introns, it is essential only for splicing the *N. crassa* mt LSU intron (Mohr et al., 2001; Chen et al., 2004). RNA footprinting experiments with the *N. crassa* mt LSU and *ND1* introns showed that CYT-18 interacts primarily with the phosphodiester backbone of group I intron RNAs on the side opposite the RNA's active-site cleft (Caprara et al., 1996a,b). In both introns, most of the putative protein protection sites were in the P4-P6 domain, with a smaller number in the P3-P9 domain. Directed hydroxyl-radical cleavage assays indicated that CYT-18's nucleotide-binding fold and C-terminal domains both contribute to the binding of group I intron (Myers et al., 2002).

The finding that CYT-18 is a TyrRS raised the possibility that it adapted to function in splicing by recognizing conserved tRNA-like structural features of group I intron RNAs (Akins and Lambowitz, 1987). This possibility was advanced by structure modeling, which suggested an extended three-dimensional overlap between tRNA^{Tyr} and group I intron catalytic core (Caprara et al., 1996a), and this overlap was used to superimpose the group I intron RNA on tRNA^{Tyr} in a *T. thermophilius*/tRNA^{Tyr} cocrystal structure (Myers et al., 2002).

Although all TyrRSs are structurally similar, only CYT-18 and the mt TyrRS of the closely related fungus *Podospora anserina* have been found to function in group I intron splicing, while the bacterial TyrRSs and the yeast mt TyrRSs do not function in splicing (Kämper et al., 1992; unpublished data). Thus, the *N. crassa* and *P. anserina* mt TyrRSs must have adaptations of the basic structure that confer splicing activity. Possible candidates for these structural adaptations include a N-terminal extension with a predicted α-helix (H0) and two small insertions (I and II), which are found in the nucleotide-binding folds of CYT-18 and the *P. anserina* mt TyrRS, but not in non-splicing TyrRSs (Figure 1.3; Mohr et al., 2001). Secondary amino acid changes in *cyt-18-1* that restore splicing but not TyrRS activity are located in H0 and insertion I (see Figure 1.3), and biochemical experiments showed that deletion of H0 strongly inhibits splicing, but not TyrRS activity with *E. coli* tRNA^{Tyr} (Cherniack et al., 1990). Directed-hydroxyl radical cleavage assays showed that insertion II lies near the P5/L9 junction in both the *N. crassa ND1* and mt LSU introns (Myers et al., 2002). Finally, the C-terminal domains of CYT-18 and the *P. anserina* mt TyrRS are substantially larger than those of the non-splicing bacterial and yeast mt TyrRS, with several additional insertions that could also contribute to RNA splicing (Figure 1.3).

Chapter 2. Crystal structure of C-terminally truncated CYT-18

X-ray crystallography of bacterial TyrRSs has been hindered by the flexibly attached C-terminal domain. Initial bacterial TyrRS structures were solved for C-terminally truncated proteins or had a completely disordered C-terminal domain (Bhat et al., 1982; Brick & Blow, 1987; Brick et al., 1989). Attempts to crystallize full-length CYT-18 in multiple laboratories were unsuccessful. Following precedents for bacterial TyrRSs, we attempted to crystallize a C-terminally truncated CYT-18 protein, denoted CYT-18/ΔC424-669. This construct contains the nucleotide-binding fold and intermediate α-helical domain of CYT-18, but lacks most of the C-terminal RNA-binding domain, as well as the N-terminal mt import sequence, which is not essential for CYT-18 function. Despite the lack of the C-terminal domain, the ΔC424-669 protein efficiently splices most group I introns, the only known exception being the *N. crassa* mt LSU intron (Mohr et al, 2001; Chen et al., 2004).

2.1 EXPRESSION AND PURIFICATION

CYT-18/Δ424-669 protein used for crystallization was expressed in *E. coli* HMS174(DE3) from the pEX602 plasmid (Mohr et al., 2001) This plasmid contains an premature stop after amino acid 423. The truncated protein has a molecular weight of 44 kDa per monomer. CYT-18/Δ424-669 was purified via the PEI-precipitation procedure (Saldanha et al., 1995). Following ammonium sulfate precipitation, the protein was dialysed against Buffer A (25 mM Tris-HCl pH 7.5; 25 mM KCl; 10% glycerol) and purified by ion-exchange chromatography using 15-S and 15-Q ion-exchange resins (Amersham-Bioscience, Piscataway, NJ). The CYT-18/Δ424-669 protein was eluted from the 15-S cation exchange column using a 25 mM to 1M KCl gradient. The protein eluted as a sharp peak at a KCl concentration of ~225 mM. Typical protein preparations were > 95% pure based on SDS-PAGE gels stained with Coomasie blue.

2.2 CRYSTALLIZATION

CYT-18/ΔC424-669 crystals were grown by sitting-drop vapor diffusion. Equal volumes of protein (7-10 mg/ml in 25 mM KCl, 25 mM Tris-HCl, pH 7.5, 10% glycerol) and precipitant solutions (200 mM $MgCl_2$, 100 mM Tris-HCl, pH 8.5, 30% PEG 4000, 1 mM L-tyrosine) were mixed and equilibrated at 22.5°C against 0.5 ml of precipitant solution. Crystals appeared overnight and grew to 400 μm x 50 μm x 10 μm within one week.

2.3 DATA COLLECTION

Data were collected for native crystals and crystals soaked in 0.5 M KI solution to provide phase information through single isomorphous replacement with anomalous scattering (SIRAS). Native crystals were harvested with a nylon loop and transferred to the precipitant solution containing 15% glycerol for one minute, then flash cooled directly in a liquid nitrogen stream at 100 K. Iodide derivatives were transferred to precipitant solution containing 15% glycerol and 0.5 M KI for one min before flash cooling. Data were collected on an in-house rotating copper anode X-ray source and MAR345 image plate detector (MAR Research, Norderstedt, Germany) and processed with HKL2000 (Otwinowski & Minor, 1997). Data collection statistics for both native and KI soaked crystals are given in Table 2.1.

2.4 STRUCTURE DETERMINATION AND REFINEMENT

The structure was solved by molecular replacement using a search model based on a monomer of the *B. stereaothermophilus* tyrosyl-tRNA synthetase (Bhat et al., 1982), with `additional `SIRAS phase information from an` iodide derivative. The monomer was

Table 2.1. X-ray crystallographic data collection and refinement

	Native	KI
Data collection		
Wavelength (Å)	1.54	1.54
Resolution (Å)	20-1.95	10-2.2
Reflections:		
Measured	152374	128815
Unique	28540	19313
Completeness (%)[a]	97.8 (89.1)	97.2 (95.4)
I / σ (I)	20.7 (3.0)	6.1 (2.0)
R_{sym} (%)	10.7 (25.4)	10.1 (34.4)
Unit cell (Å)	a = 104.88 b = 73.21 c = 56.79 $\alpha = \gamma = 90^o$ $\beta = 111.35^o$	a = 104.66 b = 73.61 c = 56.72 $\alpha = \gamma = 90^o$ $\beta = 111.32^o$
Refinement		
Resolution (Å)	20-1.95	
Number of atoms:		
Protein	2963	
Water	200	
R_{work} (%)	17.4 (16.7)	
R_{free} (%)[b]	23.5 (28.3)	
R.m.s. deviation:		
Bond length (Å)	0.018	
Bond angles (°)	1.621	
Ramachandran plot (%):		
Most favored	95.4	
Allowed	4.0	
Generously allowed	0.6	

[a] Numbers in parentheses reflect values in the highest bin (2.0-1.95 Å and 2.28-2.20 Å).

[b] 7% of the data were excluded from the refinement for calculation of R_{free}.

used for molecular replacement based on a Matthew's coefficient indicating a high probability of a single molecule in the asymmetric unit (Matthews, 1968). Rotation and translation searches were preformed with CNS (Brünger et al., 1997). The best molecular replacement solution had a correlation coefficent of 25% with R_{work} = 50.48% and R_{free} = 53.76% following rigid body refinement in CNS. Ten rounds of model building, performed with Xfit (McRee, 1992) and refinement in CNS resulted in a R_{work} = 37.8% and R_{free} = 46.8%. At this point in refinement, electron density maps for the protein core were well defined, but a number of loop regions showed weak or no density. This led to an attempt to obtain independent phase information from halide replacement by soaking in KI (Dauter et al., 2000). Twelve halide sites were found using SOLVE (Terwilliger & Berendzen, 1999), six of which were confirmed by difference Fourier analysis using phases from the partially fit MR model. SIRAS and model phases were combined for electron density map calculations and resulted in appearance of several loops and insertion regions that could not be identified by model phases alone. Final refinement statistics are given in Table 2.1

2.5 STRUCTURAL FEATURES

2.5.1 Overall structure

Figure 2.1 shows a stereo view of the CYT-18/ΔC423-669 monomer structure superimposed on that of TyrRSBst (Brick et al., 1989; PDB accession number: 3TS1). The nucleotide-binding fold of bacterial TyrRSs consists of a highly conserved Rossmann-fold with a six-member β-sheet (βA-βF) surrounded by α-helices (H1-H11), and the intermediate α-helical domain consists of 5 α helices (H1'-H5'). As expected from the high degree of sequence homology (33% identity, 52% similarity), CYT-18's nucleotide-binding fold and intermediate α-helical domains superimpose well on those of

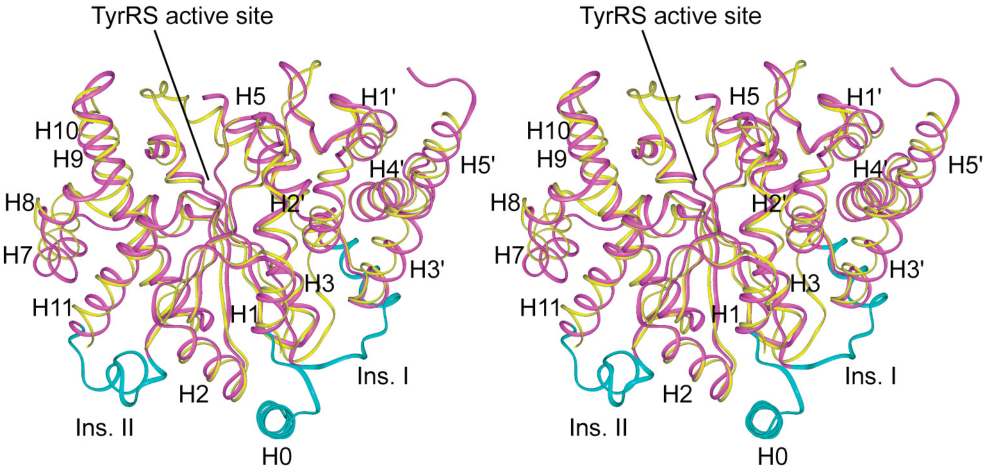

Figure 2.1. Structure of CYT-18/ΔC424-669. Stereo ribbon diagrams showing structural overlap of the nucleotide-binding fold and α-helical domains of CYT-18/ΔC424-669 (purple) and *B. stearothermophilus* TyrRS (yellow). The *B. stearothermophilus* TyrRS monomer structure (PDB ID: 3TS1) was superimposed on that of CYT-18 by least-squares fit of the Cα atoms of the nucleotide-binding fold's central β-sheet. CYT-18-specific insertions are in cyan.

the *B. stearothermophilus* TyrRS, except for the three CYT-18-specific additions: the N-terminal extension H0 and insertions I and II (cyan). The 42 Cα atoms of the central β-sheets of the two proteins have an average RMSD of 1.0 Å.

2.5.2 CYT-18-specific insertions

As indicated previously, the CYT-18-specific insertions are potentially involved in RNA splicing. Multiple sequence alignments show that H0 and insertion II are not present in non-splicing bacterial or yeast mt TyrRSs, while insertion I is not present in bacterial TyrRSs, but is partially conserved in the yeast and other fungal TyrRS (Figure 1.3 and data not shown). The X-ray crystal structure shows that the N-terminal extension and the two insertions are located on the side of the nucleotide-binding fold opposite the TyrRS active site and interact with each other directly or indirectly (Figure 2.2). As surmised previously from secondary structure predictions of wild-type CYT-18 and functional variants isolated genetically (Mohr et al., 2001), the N-terminal extension is comprised largely of an α-helix, denoted H0. The structure shows that H0 is stabilized against the protein core via hydrophobic interactions involving W52 and R55, which stack together on R95 from the loop between H2 and βB. The C-terminal portion of H0 packs through hydrophobic interactions between I58 and residues in H2, and ionic interaction between K64 and E126 (Figure 2.2). The critical nature of these interactions is indicated by the previous genetic experiments, which showed that W52 was invariant and R55 and I58 were highly conserved in splicing-competent CYT-18 variants isolated from protein libraries in which parts of H0 were partially randomized (Mohr et al., 2001). The packing of H0 against the protein core presumably establishes a specific orientation of H0 required for RNA splicing activity.

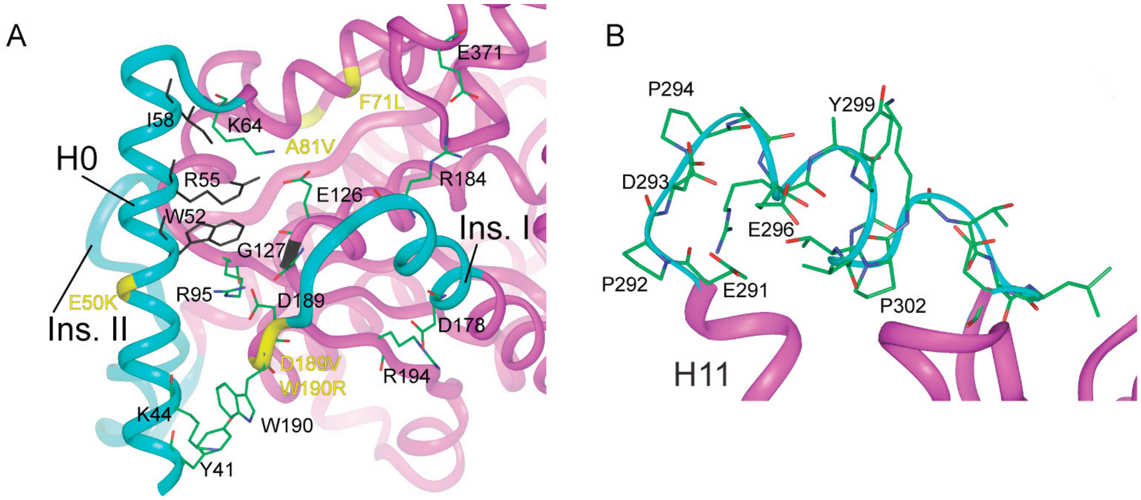

Figure 2.2. Structure of the CYT-18 N-terminal insertions. (A) Structure of the N-terminal extension. Amino acid residues that are highly conserved (>80%) in splicing-competent CYT-18 variants isolated from a library with partially randomized H0 sequences (Mohr et al., 2001) are shown with black side chains. The locations of the *cyt-18-1* mutation, G127E, and second-site mutations that restore splicing but not TyrRS activity are shown as black and yellow ribbon segments, respectively. Residues with green side chains are referred to in the text. (B) Structure of insertion II showing the location of N-terminal acidic and proline residues.

Insertion I is an extension of α-helix H5 and a portion of the following loop, and like the N-terminal extension, makes a number of contacts with the protein core. These include ionic interaction between D178 of insertion I and R194 of βD, which may stabilize the extended H5 helix against the β-sheet core, and between R184 and E371 in H3' of the intermediate α-helical domain. Additionally, there is a salt bridge between D189 of insertion I and R95, a residue that is also involved in stacking interactions with H0, as well as localized stacking interactions between W190 of insertion I and K44 and Y41 of H0. These direct and indirect H0/insertion I interactions form a pocket around the βC/H3 turn, which contains G127, the site of the *cyt-18-1* mutation (Figure 2.2).

Insertion II consists of 18 residues that are located just downstream of a cluster of basic amino acid residues in H11 that are involved in binding $\mathrm{tRNA}^{\mathrm{Tyr}}$ in bacterial TyrRSs (Bedouelle and Winter, 1986). This insertion begins with a negatively charged loop downstream of H11, contains a short non-canonical proline-capped α-helix (H12), and ends in another loop (Figure 2.2). Insertion II is of similar size in the *N. crassa* and *P. anserina* mt TyrRSs, but sequence conservation is limited to N-terminal loop and parts of the helical region (Figure 1.3). L306 and D307 in insertion II interact with H2, which also interacts with H0. Previous directed hydroxyl radical cleavage assays showed that E296 within the helical region of insertion II is close to P5 and P9 in both the *N. crassa* mt LSU and *ND1* introns (Myers et al., 2002).

Chapter 3. Biochemical characterization of CYT-18 insertion II deletion mutations

As described in Chapter 1, only CYT-18 and the mt TyrRS of the closely related fungus *P. anserina* have been found to function in group I intron splicing. These proteins have three N-terminal-domain insertions relative to the bacterial TyrRSs and have been previously identified as potentially important for group I intron splicing. Specifically, deletion of the N-terminal insertion H0 inhibits splicing activity (Mohr et al., 2001) and insertion II had been shown to be in physical proximity to the intron RNA based on hydroxyl radical cleavage assays (Myers et al., 2002). Determination of the CYT-18/ΔC424-669 structure has allowed visualization of these insertion regions for the first time. Because of the potential interaction between insertion II and the group I intron RNA, I chose to investigate the role of insertion II in group I intron splicing by constructing three small deletions in this element (Δ291-297, Δ297-303, Δ303-307) in both the full-length and CYT-18/ΔC424-669 proteins.

3.1 CONSTRUCTION OF CYT-18 INSERTION II DELETION MUTANTS

CYT-18 insertion II deletion mutants were constructed by whole plasmid single-primer PCR. The PCR primer contained ~20 complementary base pairs flanking either side of the desired mutation site with the mutation directly incorporated into the primer sequence. Following PCR, the template plasmid was digested with the restriction endonuclease *Dpn*I which is specific for the methylated template DNA. Following digestion, the mixture was transformed directly into *E. coli* DH5α, plasmids were isolated from single colonies and sequenced to verify the mutation and the integrity of the open-reading frame. Mutant proteins were expressed and purified as described in Chapter 2.1.

3.2 IN VITRO SPLICING

Unless otherwise indicated, splicing assays were carried out by incubating 50 nM ^{32}P-labeled precursor RNA and 25 nM CYT-18 protein dimer in 200 μl of 100 mM KCl, 5 mM $MgCl_2$, 20 mM Tris-HCl, pH 7.5, 5 mM DTT, 10% glycerol, 1 mM GTP, and 0.1 mg/ml bovine serum albumin. The RNA was preincubated at 37° C for 5 min prior to adding CYT-18. At different times, 8 μl portions were removed, and the reaction was terminated by adding EDTA to 50 mM, followed by extraction with phenol-chloroform-isoamyl alcohol (25:24:1; phenol-CIA). Products were analyzed in denaturing 6% polyacrylamide gels, which were dried and quantified with a PhosphorImager.

3.3 TYROSYL ADENYLATION AND TYRRS ACTIVITY ASSAYS

Tyrosyl-adenylation assays were as described (Mohr et al., 2001) with initial protein concentrations determined by A_{280}. TyrRS assays were with 50 nM CYT-18 and 7.4 μM *E. coli* $tRNA^{Tyr}$ (Sigma, St. Louis, MO) or 8.4 μM *N. crassa* mt $tRNA^{Tyr}$ transcript in 20 μl of 100 mM KCl, 15 mM $MgCl_2$, 50 mM Tris-HCl, pH 7.5, 5 mM ATP and 1 μCi L-[3,5-^{3}H]tyrosine (56 Ci $mmol^{-1}$; Amersham Bioscience). Reactions were initiated by adding protein, incubated at 37° C for 5 min, a time determined to be in the linear range, and terminated by precipitation with 1 ml 10% (w/v) tricholoroacetic acid. The precipitates were collected on a glass-fiber filter and counted in scintillation cocktail with a Beckman LS 6500 scintillation counter.

3.4 ACTIVITIES OF CYT-18 INSERTION II DELETION MUTANTS

All three of the insertion II deletions had little if any effect on tyrosyl-adenylation activity of either the full-length CYT-18 or CYT-18/ΔC424-669 proteins, or on the TyrRS activity of the full-length protein with either *E. coli* $tRNA^{Tyr}$ or *in vitro* transcribed *N. crassa* mt $tRNA^{Tyr}$ (Figure 3.1). TyrRS activity of CYT-18/ΔC423-669 could not be assayed because the C-terminal domain is required for tRNA charging (Carter et al., 1986). By contrast, all three insertion II deletions inhibited the splicing of the *N. crassa ND1* intron, with this inhibition being much more pronounced in the C-terminally truncated protein (Figure 3.1). The specific inhibition of splicing by the insertion II mutants may reflect loss of direct contacts between insertion II and the intron RNA, but could also result from structural rearrangements caused by the deletions. The less severe effects in the full-length protein suggest that interactions with CYT-18's C-terminal domain can partially compensate for the loss of insertion II interactions.

3.5. STRUCTURE DETERMINATION OF CYT-18/ΔC424-669 INSERTION II DELETION MUTANTS

To determine if structural rearrangements in the insertion II deletion mutants were responsible for the inhibition of splicing, I crystallized and determined the structure of the Δ291-297 and Δ297-303 C-terminally truncated proteins. These mutant proteins were prepared in the same manner as in Chapter 2.1. The proteins crystallized using the same conditions as the CYT-18/ΔC424-669 protein described in Chapter 2.2. Surprisingly, the Δ291-297 mutant crystallized in a different space group than the CYT-18/ΔC424-669 protein, having a dimer in the asymmetric unit (Table 3.1). The structures were determined by molecular replacement using the CYT-18/ΔC424-669 structure as a starting model. The insertion II region was disordered in the Δ297-303 proteins, but was present in both monomers of the Δ291-297 protein (Figure 3.2). The disordered loop between H4 and H5 in the CYT-18/ΔC424-669 structure was present in both the Δ291-297 and Δ297-303 structures. Importantly, these structures show that there are no significant structural changes associated with the insertion II deletion mutants outside the local structure of the insertion (Figure 3.2).

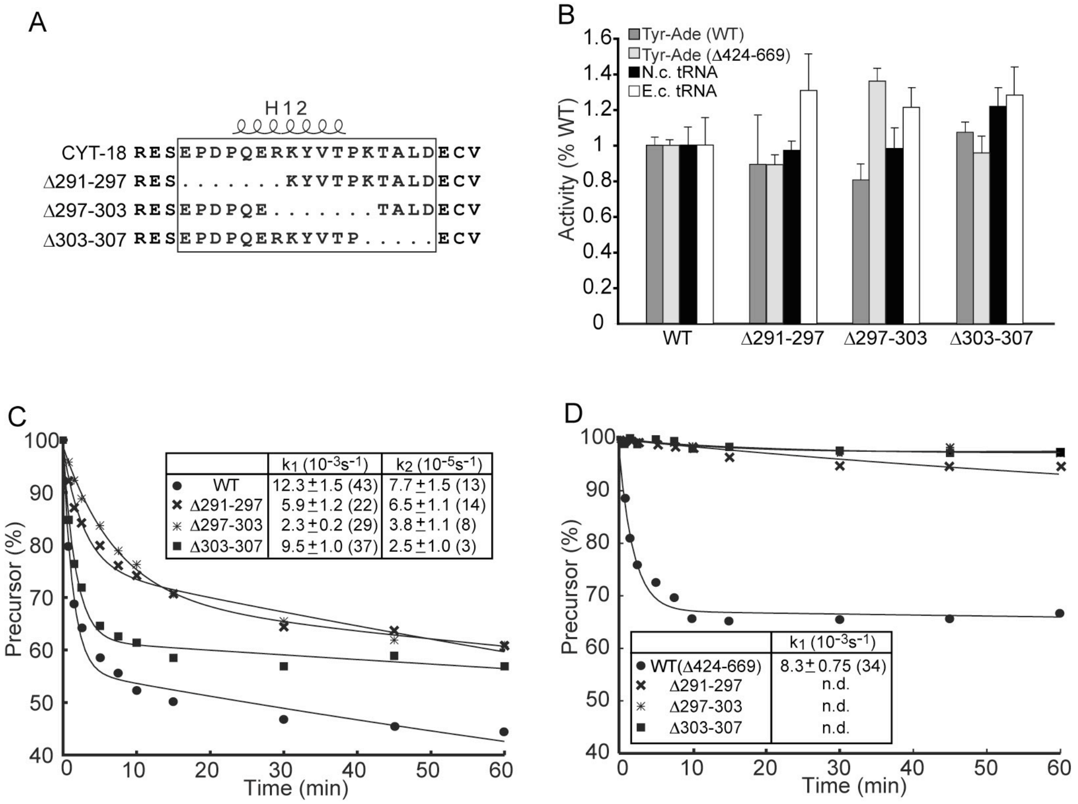

Figure 3.1. Biochemical analysis of CYT-18 mutants with small deletions in insertion II. (A) Amino acid sequence of insertion II mutations. (B) Tyrosyl-adenylation and TyrRS assays. Activities are expressed as percent that of wild-type CYT-18 or CYT-18/Δ424–669, with bars indicating S.D.s for three determinations. (C and D) *In vitro* splicing assays. Splicing time courses were carried out with 50 nM *ND1* RNA and 25 nM CYT-18 (C) or 1 μM *ND1* RNA and 500 nM CYT-18/Δ424–669 (D). The plots show the best fit of the data to equations with one or two exponentials. The boxes show k_{obs} for the fast (k_1) and slow (k_2) phases, with the numbers in parentheses indicating the proportion of precursor RNA spliced in each phase. Similar results were obtained for three different preparations of each protein.

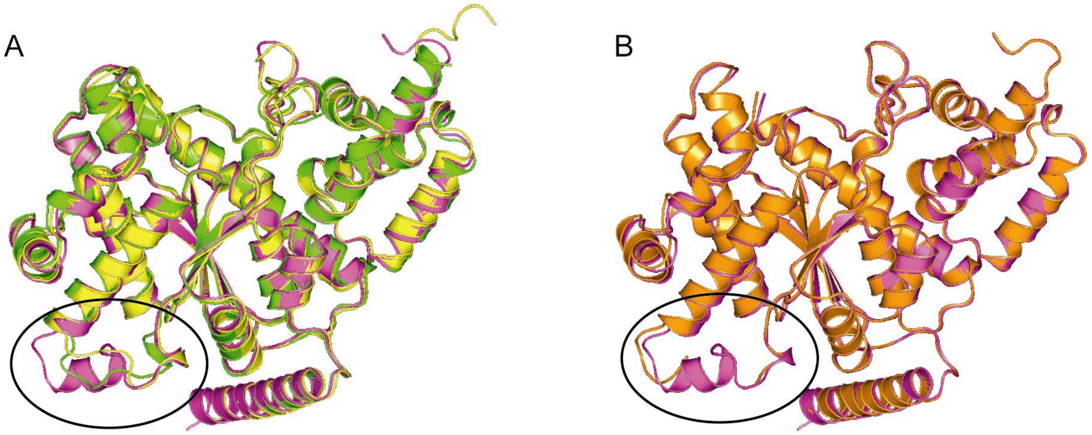

Figure 3.2. Crystal structures of CYT-18 insertion II deletion mutants. (A) A ribbon diagram of the two subunits of the Δ291-297 protein (yellow and green) superimposed with the CYT-18/Δ424−669 structure (magenta) by least-squares fitting. The only significant structural differences between the mutant protein and CYT-18/Δ424−669 occur at the site of the insertion, highlighted in the black oval. (B) A ribbon diagram of the Δ297-303 structure (orange) superimposed on the CYT-18/Δ424−669 structure (magenta). Insertion II is disordered in this structure, while the remainder of the proteins are essentially identical.

Table 3.1. X-ray crystallographic data collection and refinement for insertion II mutants

	Δ291-297	Δ297-303
Data Collection		
Space group	$P2_1$	C2
Unit cell (Å)	a = 77.55	a = 106.12
	b = 64.02	b = 72.79
	c = 86.59	c = 57.61
	$\alpha = \gamma = 90^o$	$\alpha = \gamma = 90^o$
	$\beta = 104.71^o$	$\beta = 111.05^o$
Wavelength (Å)	1.54	1.54
Resolution (Å)	20.0-2.23	20.0-2.26
Reflections:		
Measured	68481	30596
Unique	30426	15288
Completeness (%)[a]	97.3 (83.9)	92.5 (88.3)
I / _ (I)	8 (1.8)	17 (2.3)
R_{sym} (%)	8.3 (35.3)	3.8 (29.6)
Refinement		
Resolution	14.0-2.23	14.0-2.30
Monomers in asymmetric unit	2	1
Number of atoms:		
Protein	5822	2886
Water	181	158
R_{work} (%)	17.9 (22.0)	21.4 (21.5)
R_{free} (%)[b]	26.5 (37.2)	33.4 (34.5)
R.m.s. deviation:		
Bond length (Å)	0.025	0.034
Bond angles (°)	2.217	2.709

[a] Numbers in parentheses reflect values in the highest bin.

[b] 5% of the data were excluded from the refinement for calculation of R_{free}.

Chapter 4. Biochemical characterization of the CYT-18/*ND1* intron interaction

Directed hydroxyl-radical cleavage assays provide proximity information from a known position in the protein to nucleotide positions in the RNA. EPD-Fe conjugated at sites of single-cysteine substitutions is used to generate hydroxyl radicals that cleave the sugar-phosphate backbone of bound RNA (Hall and Fox, 1999). EPD-Fe has a ~ 14 Å tether and the radical can travel ~ 10 Å in solution to provide a distance constraint of ~ 24 Å from the amino acid to the nucleotide sugar in the event of a positive cleavage. In previous work, EPD-Fe conjugated at four sites in CYT-18's nucleotide-binding fold (R213C, R214C, A287C, E296C) cleaved both the *N. crassa ND1* and mt LSU introns in P4, P5, and P9, while EPD-Fe conjugated at sites in the C-terminal domain cleaved the *ND1* intron in P6-J6/6a, P3, and P8, and the mt LSU intron more extensively in P4, P6, P5, and P2 (Myers et al., 2002). These cleavages were consistent with the intron RNAs being in the same orientation as tRNA^{Tyr} in the *T. thermophilus* TyrRS/tRNA^{Tyr} cocrystal structure, but were insufficient to independently define the orientation of the intron on the protein (Myers et al., 2002).

4.1 SITE-DIRECTED HYDROXYL RADICAL CLEAVAGE AND CHARACTERIZATION OF SINGLE CYSTEINE MUTANTS.

To obtain more comprehensive data, I carried out directed hydroxyl-radical cleavage assays with EPD-Fe conjugated via single-cysteine substitutions at 20 additional sites in the full-length CYT-18 protein, now using the crystal structure to systematically select surface residues in the nucleotide-binding fold and alpha-helical domains. Single-cysteine mutants were constructed by single primer PCR of whole plasmids as described in Chapter 3.1 and purified as described in Chapter 2.2. Seven new sites in the nucleotide-binding fold gave specific cleavages of the bound *ND1* intron RNA (Y41C, H167C, R182C, K193C, H201C, K205C, Y260C). Figure 4.1 shows the locations of the cleavage sites in the *ND1* intron, as well as with the cleavages obtained from the four previously identified sites in the nucleotide-binding fold (Myers et al., 2002).

The seven new single-cysteine substitution mutants that gave specific cleavages all retained wild-type tyrosyl adenylation and TyrRS activity with *E. coli* tRNA^{Tyr} both before and after EPD-Fe conjugation (data not shown). However, their splicing activities with the *ND1* intron varied. Relative to wild-type CYT-18, splicing activities before EPD-Fe conjugation were 80% for H167C, <10% for K205C, and 30-60% for the remaining mutants, and splicing activities after EPD-Fe conjugation were ~25% for Y41C, H167C, and Y260C, and < 5% for the remaining mutants (assayed by the extent of splicing after 10 minutes; data not shown). Despite the reduced splicing activities, all the EPD-Fe-conjugated proteins gave specific and reproducible intron RNA cleavages, which were not competed by a five-fold excess of non-specific competitor RNA (Figure 4.1). The remaining 13 mutants in CYT-18's nucleotide-binding fold (T83C, Y168C, R220C, S241C, K285C, T301C, S321C) and α-helical domain (Y338C, Q339C, Y348C, H413C, A416C, Y417C) retained tyrosyl adenylation and splicing activity prior to EPD-Fe modification, but gave no specific cleavages in the intron RNA. The results indicate that many of the newly identified positions that give specific cleavages are critical for splicing activity.

Figure 4.2 shows the CYT-18/ΔC424-669 structure with sites of EPD-Fe conjugation that did or did not give specific cleavages in the *ND1* intron RNA indicated in yellow and black, respectively. As a reference point, tRNA^{Tyr} is docked to the structure in the same position as in the *T. thermophilus* cocrystal structure (Yaremchuk et al. 2002).

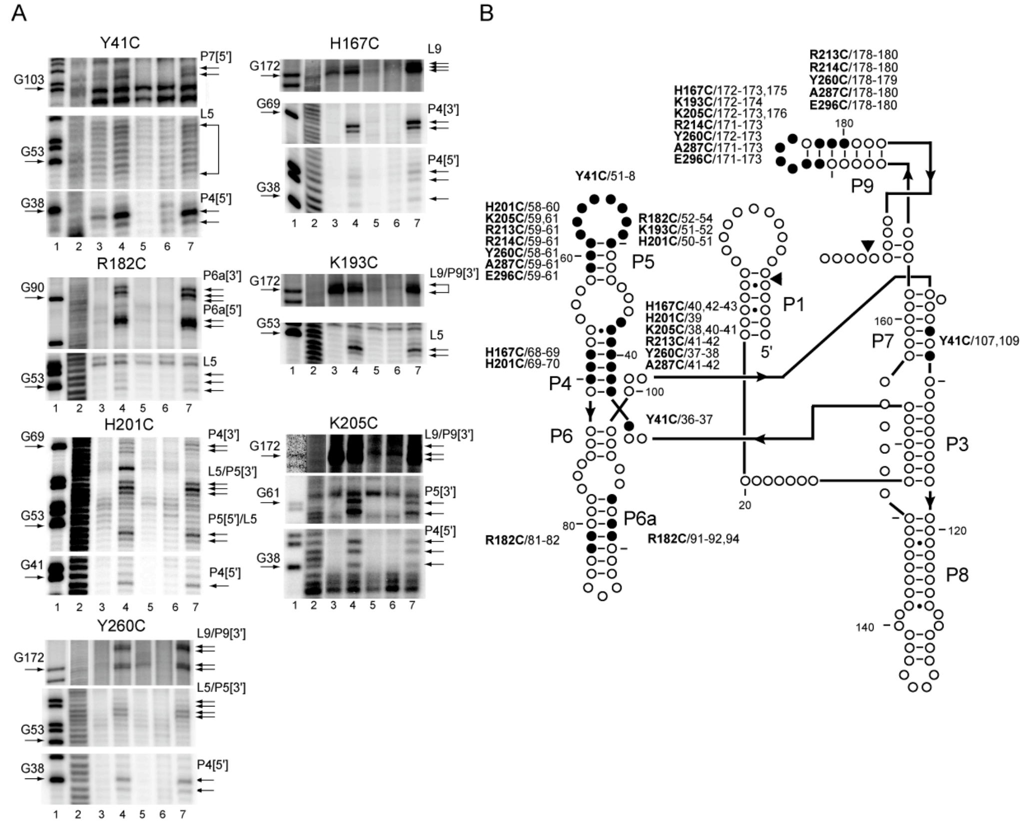

Figure 4.1. Site-directed hydroxyl-radical cleavage assays. (A) CYT-18 proteins containing EPD-Fe conjugated at sites of single cysteine substitutions were complexed with 5'-labeled *ND1* RNA, and cleavage sites were mapped by electrophoresis in denaturing polyacrylamide gels. Lanes 1 and 2, ladders generated from the 5'-labeled *ND1* RNA by partial RNase T1 and alkaline hydrolysis, respectively; lanes 3, EPD-Fe modified protein incubated with RNA in the absence of ascorbate; lanes 4-7, cleavage reactions after the addition of ascorbate (lanes 4); in the presence of 2-fold molar excess of unmodified wild-type CYT-18 protein (lanes 5); in 0.5 M KCl to disrupt complexes (lanes 6); and in the presence of 5-fold molar excess of nonspecific competitor RNA (pTRSE15/NsiI; see Myers et al., 2002) (lanes 7). (B) *ND1* intron secondary-structure diagram showing sites of nucleotide cleavage (filled) from EPD-Fe conjugated at the indicated positions in this study and Myers et al. (2002). SS, splice site.

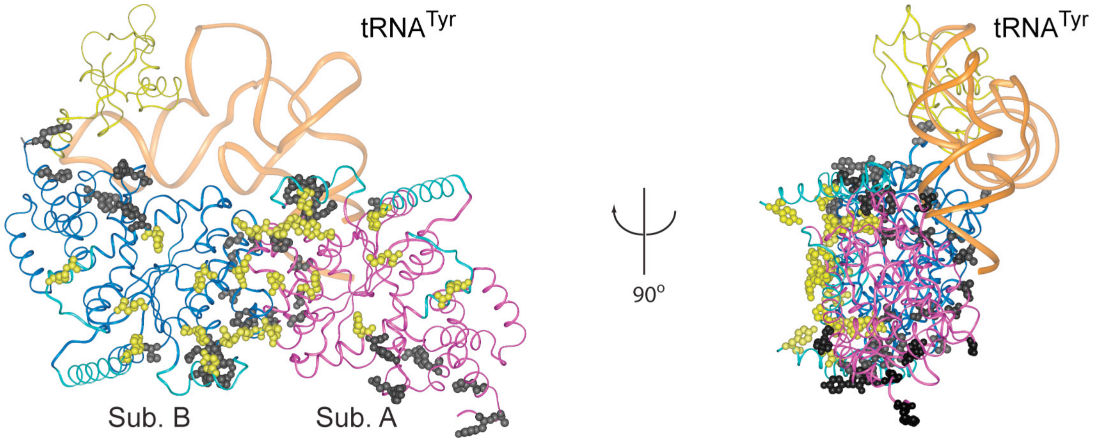

Figure 4.2. Structural summary of site-directed cleavage data. Dimeric CYT-18/ΔC424-669 with the *T. thermophilus* tRNA^{Tyr} (orange) docked as in the *T. thermophilus* TyrRS/tRNA^{Tyr} cocrystal structure (Yaremchuk et al. 2002; PDB ID: 1H3E). Subunits A (Sub. A; magenta) and B (Sub. B; blue) are defined as those that bind the tRNA acceptor and anticodon arms, respectively. Side chains at positions that did or did not give specific EPD-Fe-induced cleavages in the *ND1* intron are shown in space-filling representation colored yellow and black, respectively

All seven new positions that gave specific cleavage are on the side of the nucleotide-binding fold opposite that which binds $tRNA^{Tyr}$. Y41 is in H0 with its side chain extending into a cleft between H0 and insertion II (see Figure 2.2); H167 is in H5; R182 is at the C-terminus of the insertion I helix; K193 is at the N-terminus of βD; Y260 is at the C-terminus of helix H10; and H201 and K205 are in H6 near the dimerization interface. Additionally, the structure reveals that all four previously identified nucleotide-binding fold sites that gave specific cleavages in the *ND1* intron (Myers et al., 2002) have their side chains oriented away from the aminoacylation active site. R213 and R214 are in H7; A287 is in H11; and E296 is in the α-helical region of insertion II. Significantly, R297, the residue adjacent to E296 in insertion II, has its side chain facing toward the TyrRS active site and EPD-Fe conjugated at this position did not give specific cleavages in the intron RNA (Myers et al. 2002). The conjugated EPD-Fe positions that did or did not give RNA cleavages demonstrate that the intron RNA binds to the nucleotide-binding fold on the side opposite that which binds $tRNA^{Tyr}$.

4.2 COMPUTER-ASSISTED INTRON DOCKING

The site-directed cleavage data combined with the CYT-18/ΔC424-669 structure enabled me to model a protein-bound group I intron RNA, using the distance constraints between the cleaved nucleotide residues and the mutant cysteine positions. I chose a computer-assisted approach to this docking experiment to eliminate potential bias, fully exploit the large amount of data generated in the cleavage experiments, and provide a method to identify all possible intron orientations consistent with the distance constraint data.

A previous structural model of the *N. crassa ND1* intron (Caprara et al. 1996a) was updated based on the X-ray crystal structure of the *Azoarcus* sp. BH72 group I intron

(Adams et al., 2004). Large portions of the *ND1* model were fitted onto the *Azoarcus* structure using least-squares fitting, with additional features fitted manually. The final structure was energy minimized using CNS with base-pairing restraints set to maintain RNA secondary structure.

Docking was performed by rigid-body minimization in CNS with distance constraint information assigned as NOE distances. The group I intron was randomly rotated and translated 100 Å away from the protein dimer center-of-mass, and the intron was docked by minimizing NOE, van der Waal's, and electrostatic energy. This process was repeated 1000 times with the intron RNA randomly rotated and translated between each cycle.

Because CYT-18 is a dimeric protein, some ambiguity existed as to which protein subunit was responsible for specific cleavages, particularly in cases where the mutant cysteine positions were near the dimerization interface. Cleavages from the positions further from the dimerization interface could be assigned unambiguously, and provided initial starting point for the assignments. With these distance constraints assigned, the remaining constraints could be assigned based on whether the RNA cleavage occurred on either side of the J4/6 region of the intron.

To confirm the accuracy of the subunit assignments and examine any other possible intron orientations, a second experiment was performed in which the chain identifier in the NOE assignments for the mutant cysteine positions was randomly assigned to either of the two monomers between each cycle of rigid-body docking for each cleavage. In the latter experiment, only one central cleavage was assigned for stretches of two or more cleaved nucleotide residues to decrease the complexity of the calculation. Even with the decreased complexity there were over 8 million possible combinations, of which 200,000 were tested.

Docking from 1000 random intron starting positions with assigned cleavages resulted in 14 models that had negative van der Waal's energy and satisfied the distance constraints (4-25 Å with a mean ± S.D. of 17.4 ± 5.6 Å between Cβ at the mutant cysteine position and ribose C4' of the cleaved nucleotide residue). The ensemble of models is shown in Figure 4.3. The models had a global RMSD from each other of 6.39 ± 3.22 Å and an average RMSD of 5.05 Å from a mean model calculated with MOLMOL (Koradi et al., 1996). RMSD analysis of the 14 models by nucleotide position shows that residues from P4, P6, P3 and J8/7 have the least deviation between models (Figure 4.3). The docking experiment using the randomized subunit assignments resulted in 7 models that met the docking criteria. Of these, five were essentially identical to the models using the assigned constraints, while the other two models were rotated 180° about the protein dyad axis, but were structurally equivalent (see Chapter 4.3 and Chapter 5.1).

4.3 CYT-18/ND1 INTERACTION MODEL

In the models, subunits A and B are defined as those that bind the tRNA acceptor stem and anticodon loop, respectively. In all of the models, the intron RNA binds to the nucleotide-binding fold across the dimer interface on the side opposite the TyrRS active site. Because the intron binds on a two-fold symmetrical face of the protein dimer, it could in principle bind equally well in the opposite orientation (*i.e.*, rotated 180° about the dimer dyad axis). In the orientation shown in Figure 4.4, the minor groove of P4 runs along βD and H5 of subunit B, both of which have a number of polar and charged side chains poised to contact the RNA (H167, K171, E175, K193, R194). H0 and insertion I from both monomers have the potential to contact the intron. H0 and insertion I of subunit A are near L5, while H0 and insertion I from subunit B are near P4, P6, P6a, and P7. In all 14 models, insertion II from subunit B contacts P9, and in 7 of these models, the N-terminal loop and insertion II α-helix are positioned directly in the P9 minor

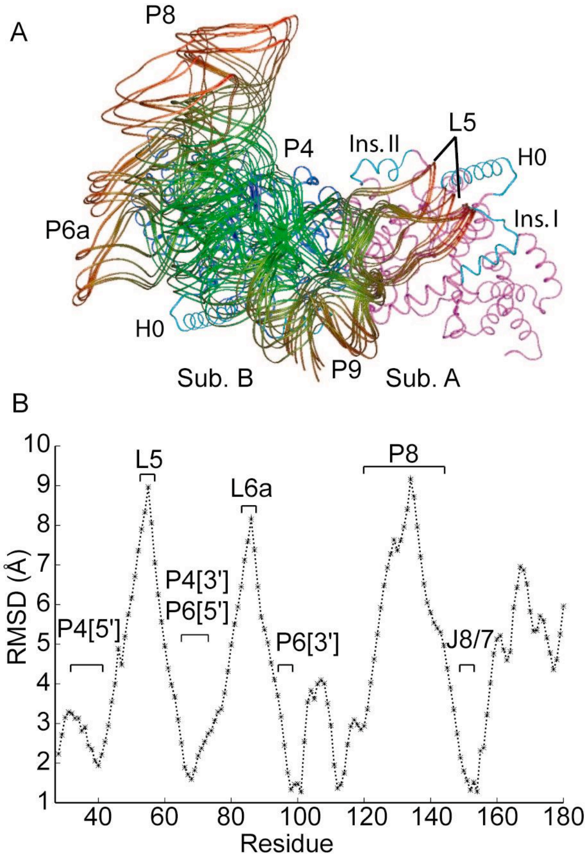

Figure 4.3. Ensemble of models of the *ND1* intron docked to CYT-18/ΔC424-669 dimer. (A) The figure shows the 14 models that have negative van der Waal's energy and satisfy the distance constraints from EPD-Fe cleavage data from 1000 random starting positions. Protein subunits A and B are purple and blue, respectively. Variations in color of the intron ribbon indicate RMSD by position with green being the lowest and red being the highest. (B) RMSD analysis of the 14 intron orientations by position determined by MOLMOL (Koradi et al., 1996).

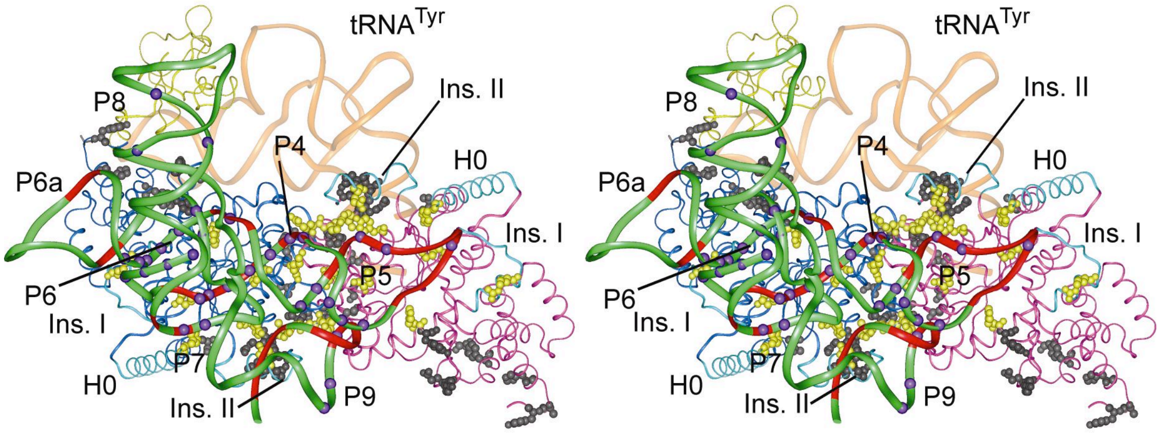

Figure 4.4. Stereo view of dimeric CYT-18/ΔC424-669 with docked *ND1* intron RNA. The model is based on optimized fit to directed hydroxyl-radical cleavage data summarized in Figure 4B. The *ND1* intron RNA (residues 27-182) is shown as a green ribbon, with purple balls indicating phosphate-backbone protections from full-length CYT-18 protein (Caprara et. al 1996a), and red ribbon segments indicating EPD-Fe-cleavage sites. The C-terminal domain of *T. thermophilus* TyrRS (yellow) is shown positioned on subunit B as in the *T. thermophilus*/tRNA$^{\text{Tyr}}$ cocrystal structure (Yaremchuk et al., 2002).

groove. In 3 of the 14 models, insertion II from subunit A is in proximity to L5, while in the remaining models the intron RNA is shifted away from insertion II of subunit A, toward H0 and insertion I (Figure 4.3).

CYT-18 interacts primarily with the phosphodiester backbone of group I intron RNAs, and as expected for electrostatic interactions, group I intron binding and splicing are strongly inhibited at elevated KCl concentrations (≥200 mM) (Caprara et al., 2001). Putative protein protection sites on the phosphodiester backbone of the bound *ND1* intron were identified previously by iodine-footprinting (Caprara et al., 1996a). Our structural models account for all of the putative protein protections in P4, P5, P7, and P9, but do not account for several protections in P3, P6 and P8 (Figure 4.4). The latter likely result from binding of the C-terminal domain, or from conformational changes in the intron RNA upon binding.

Electrostatic surface potential calculations show that the most positively charged regions of CYT-18/ΔC424-669 are on the side opposite the TyrRS active site localized around H0 and insertion I, and extend inward toward the dimerization interface along βD and H5 (Figure 4.5). Comparable positively charged regions are not found in the *B. stearothermophilus* and *T. thermophilus* TyrRSs. H0 and insertion I contribute many of the basic residues to these positively charged regions, but surprisingly, a significant number also come from basic amino acid substitutions in neighboring parts of the nucleotide-binding fold relative to bacterial TyrRSs. The latter include K171 in H5 and R192, K193, R194 in βD. Thus, not only the insertions, but also concerted amino acid changes in surrounding regions appear to have contributed to the formation of the group I intron RNA-binding site.

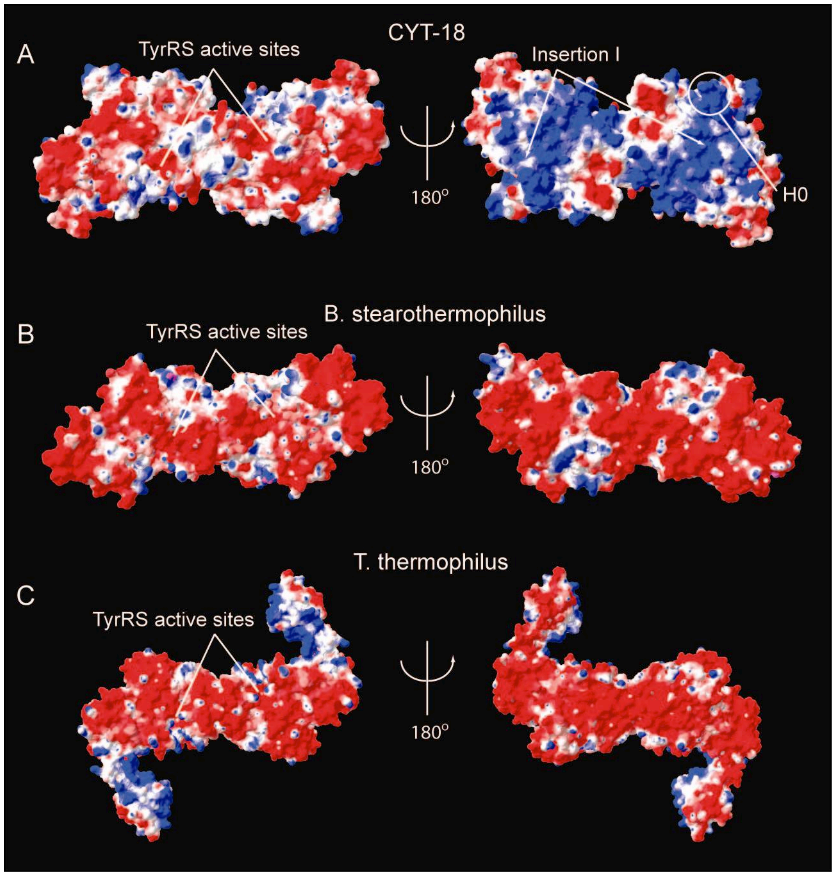

Figure 4.5. Electrostatic surface potentials of CYT-18/ΔC424-669 and bacterial TyrRSs. (A) CYT-18/ΔC424-669 dimer structure shown in space-filling representation with calculated electrostatic surface potentials. The left shows the side of the protein with the TyrRS active site, and the right shows the opposite side. Positively charged regions are shown in blue, neutral in white, and negatively charged in red. (B and C) show the same views of the *B. stearothermophilus* TyrRS (PDB ID: 4TS1) and *T. thermophilus* TyrRS (PDB ID: 1H3E).

Chapter 5. Conclusions

I determined a 1.95 Å X-ray crystal structure of a C-terminally truncated CYT-18 protein and used it in combination with directed hydroxyl radical cleavage and other biochemical assays to construct highly constrained models of CYT-18 with a bound group I intron RNA. My results indicate that the group I intron RNA, like tRNA^{Tyr}, binds across the two subunits of the homodimer, but interacts with CYT-18's nucleotide-binding fold on the side opposite the TyrRSs active site, contrary to previous hypotheses. This new group I intron RNA-binding site appears to have been created by three CYT-18-specific insertions along with concerted changes in neighboring amino acid residues.

5.1 THE ROLE OF CYT-18-SPECIFIC INSERTIONS

Several lines of evidence indicate that the region of the nucleotide-binding fold containing the insertions plays a critical role in RNA splicing. First, this region is the location of the *cyt-18-1* mutation, which inhibits both splicing and TyrRS activities, as well as second-site suppressor mutations, which restore splicing, but not TyrRS activity (Cherniack et al., 1990). Second, deletion of H0 inhibits the splicing of the *N. crassa* mt LSU intron and abolishes splicing of the *ND1* intron, without affecting TyrRS activity assayed with *E. coli* tRNA^{Tyr} (Cherniack et al., 1990; Mohr et al., 2001). Third, we find that small deletions in insertion II, which do not produce gross structural changes elsewhere, as well as many of the single cysteine-substitutions used for EPD-Fe conjugation in H0, insertion I, and neighboring residues specifically inhibit splicing without significantly affecting tyrosyl-adenylation or TyrRS activity. Finally, EPD-Fe conjugated at sites in H0, insertion I, and insertion II cleaves the sugar-phosphate backbone of bound group I intron RNA. This finding provides direct evidence for physical proximity.

The crystal structure shows that the site of the *cyt-18-1* mutation (G127E) is located away from the protein surface in a loop wedged between H0 and insertion I, where it is unlikely to interact directly with RNA (Figure 2.2). Modeling suggests that the larger glutamic acid side chain of the mutant protein has the potential for steric clashes with R95, which stacks with W52 of H0, and with D189 in insertion I, thus directly affecting key elements required for splicing. Further, the phi-psi angle of G127 is in the unallowed region for a non-glycine, and glycine is highly conserved at this position in all TyrRSs, reflecting its importance in facilitating the sharp turn between H3 and βC. Disruption of this turn by the G127E mutation may affect the organization of the β-sheet that forms part of the tyrosine-binding site. The second-site suppressor mutations that restore splicing but not TyrRS activity are clustered in H0, H1, βA, and insertion I and presumably restore the local structure required for RNA splicing, but not the β-sheet structure required for TyrRS activity.

Combined with the structure, the EPD-Fe cleavage data show unambiguously that the group I intron RNA binds to the nucleotide-binding fold across the surface of the two subunits of the homodimer on the side opposite the TyrRS active site. In the models shown in Figures 4.3 and 4.4, the P4-P6 stacked helices interact predominantly with subunit B, with the minor groove of P4 running along βD and H5, and the major groove of P6 lying near insertion I. All three CYT-18-specific insertions in subunit B potentially interact with the intron RNA: H0 with P4 and P7, insertion I with P6, J6/6a, and P6a, and insertion II with P9. In subunit A, H6 potentially interacts with P5, while either H0/insertion I or insertion II may interact with L5. Thus, in addition to the importance of the insertions, the models readily explain how CYT-18 could stabilize the structure around the junction of the P4-P6 stacked helices, as well as the critical L9/P5 tetraloop-receptor interaction between the group I intron's P4-P6 and P3-P9 domains, in agreement with a large amount of previous biochemical and genetic data. As indicated previously, because the intron-binding site is on a two-fold symmetrical face of the protein in the crystal structure, the intron could in principle bind equally well in the orientation opposite that shown in Figure 4.4. However, in solution or intracellularly the structures of the two monomers may differ sufficiently to define a preferred orientation for intron binding.

5.2 THE ROLE OF CYT-18'S C-TERMINAL DOMAIN

Although CYT-18's C-terminal domain is not required for the splicing the *N. crassa ND1* intron and most other group I introns tested, it nevertheless contributes to the binding of these introns. The C-terminally truncated protein has an ~5-fold higher k_{off} for the *ND1* intron than does the wild-type protein (Mohr et al., 2001), and genetic experiments show that the C-terminal domain is essential for splicing mutant group I introns with certain structural defects (Chen et al., 2004). Previous experiments with the full-length CYT-18 showed that EPD-Fe conjugated at sites in the C-terminal domain cleaves the *ND1* intron in P8, P3, and P6-J6/6a, and the mt LSU intron more extensively in P2, P6a, P6, P4, and P5 (Myers et al., 2002). Examination of the models shows that most of these cleavages can be explained if the flexibly linked C-terminal domain, which is larger in CYT-18 than in bacterial TyrRSs, shifts from the position used to bind $tRNA^{Tyr}$ to interact with the corresponding regions of the intron RNA. The C-terminal domain cleavages in P5 in the mt LSU are distant from the others and could reflect interaction with the C-terminal domain of the second monomer. The predicted binding sites for the C-terminal domain show how it could help stabilize interactions between the two major helical domains of the group I intron catalytic core, in agreement with the previous genetic data showing that the C-terminal domain is necessary to suppress mutations affecting the critical interdomain L9-P5 and L2-P8 tetraloop-receptor interactions (Chen et al., 2004). The loss of C-terminal domain interactions likely explains the finding that deletions in insertion II, which may also help stabilize the L9-P5 interaction, have much greater effects in CYT-18/Δ424-669 than in the full-length protein (see Chapter 3).

CYT-18 functions in splicing by recognizing a conserved a three-dimensional structure of group I introns (Caprara et al., 1996a). The binding site identified here for the *ND1* intron presumably accommodates other group I introns. Nevertheless, there appear to be significant differences in the binding of different introns. The deletion of H0, which the model suggests interacts directly with the *ND1* intron, abolishes the binding and splicing of that intron (Mohr et al., 2001). However, deletion of H0 decreases the binding affinity for the mt LSU intron by only about three-fold, and the residual active protein splices the mt LSU intron almost as well as the wild-type protein (Mohr et al., 2001). Conversely, the CYT-18/Δ424-669 protein does not stably bind the mt LSU intron (Mohr et al., 2001), indicating that in the absence of the C-terminal domain interactions, the N-terminal binding site identified here does not contribute substantially to the binding energy for the mt LSU intron.

Surprisingly, tRNA-like structural similarities appear to play no role in the interaction of the group I intron RNA with the nucleotide-binding fold. The group I intron's P4-P6 stacked helices and the tRNA's D-anticodon arm stacked helices, which are structural cognates (Michel and Westhof, 1990; Caprara et al., 1996b), both bind roughly parallel to the long axis of the protein dimer, but at non-overlapping sites separated by more than 90° (Figure 4.3). It remains possible that tRNA-like structural similarities play a role in the interaction with the C-terminal domain, since EPD-Fe conjugated at or near C-terminal domain positions that interact with $tRNA^{Tyr}$ gave specific cleavages in both the *ND1* and mt LSU introns (Myers et al., 2002). As expected, the mt LSU intron, which interacts strongly with the C-terminal domain, is a potent competitive inhibitor of aminoacylation (Guo et al., 1992), while the *ND1* intron, which interacts more weakly with the C-terminal domain, is a correspondingly weaker inhibitor of aminoacylation (data not shown).

5.3 EVOLUTION OF PROTEIN-DEPENDENT SPLICING

Although CYT-18's nucleotide binding fold contains a tRNA-binding site and group I introns contain some structural features similar to those of tRNAs (Caprara et al., 1996b), the tRNA-binding site is not used for binding the group I intron RNA. A likely explanation is that the tRNA-binding site in the nucleotide-binding fold could not be distorted to accommodate a group I intron RNA without disrupting its critical function in aminoacylation. The *Aspergillus niger* group I intron-encoded protein I-*AniI*, a homing endonuclease that adapted to function in splicing, also appears to have evolved a distinct binding site for the intron RNA that differs from that for its original DNA substrate (Bolduc et al., 2003). Similarly, the splicing activity of the *Saccharomyces cerevisiae* mt LeuRS, which functions together with a maturase to splice the group I introns aI4 and bI4, appears to be associated with the CP1 domain, a nucleotide-binding fold insertion, which functions in amino acid editing (Rho et al., 2002).

The finding that only the *N. crassa* and *P. anserina* mt TyrRSs and not bacterial or yeast mt TyrRSs function in splicing group I introns suggests that the protein adapted to function in splicing after the divergence of *N. crassa* and *P. anserina* from yeast (Kämper et al., 1992). The mt TyrRS may have been amenable to this adaptation because its elongated homodimer structure could accommodate the relatively large intron RNAs and because the flexibly attached C-terminal domain could transition between binding the intron RNA or tRNA on different sides of the protein. The ability of CYT-18 to promote the splicing of multiple group I introns may have been acquired in several steps. Initially, the N-terminal extension, which could have evolved from the mt targeting sequence (Mohr et al., 2001), and/or insertion I, which is present in other mt TyrRSs (see Figure 1), may have functioned to splice a single intron, perhaps inefficiently. Subsequent acquisition of insertion II, mutations of neighboring amino acid residues in the nucleotide-binding fold, and C-terminal domain insertions may have led to more efficient splicing of additional introns. Once begun, this process would lead progressively to a common splicing apparatus for multiple introns because (i) the binding site for one group I intron could accommodate other introns with similar structures and (ii) association of the intron RNAs with the protein provides opportunities for additional interactions leading to increased protein dependence. A similar process dictated by the evolution of common binding sites and progressively increasing protein dependence may have been involved in the evolution of group II introns into nuclear pre-mRNA introns. With the structural information now available, mt TyrRSs should provide an excellent model system for studying both the evolution of splicing activity and how essential proteins can progressively acquire new functions.

References

Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J., and Strobel, S.A. (2004). Crystal structure of a self-splicing group I intron with both exons. Nature *430*, 45-50.

Akins, R.A., and Lambowitz, A.M. (1987). A protein required for splicing group I introns in Neurospora mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof. Cell *50*, 331-345.

Batey, R.T., Rambo, R.P., & Doudna, J.A. (1999). Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Edn Engl* **38**, 2326-2343.

Bedouelle, H., and Winter, G. (1986). A model of synthetase/transfer RNA interaction as deduced by protein engineering. Nature *320*, 371-373.

Bolduc, J.M., Spiegel, P.C., Chatterjee, P., Brady, K.L., Downing, M.E., Caprara, M.G., Waring, R.B., and Stoddard, B.L. (2003). Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. Genes Dev. *17*, 2875-2888.

Brick, P., Bhat, T.N., and Blow, D.M. (1989). Structure of tyrosyl-tRNA synthetase refined at 2.3 Å resolution. Interaction of the enzyme with the tyrosyl adenylate intermediate. J. Mol. Biol. *208*, 83-98.

Brick, P., and Blow, D.M. (1987). Crystal structure of a deletion mutant of a tyrosyl-tRNA synthetase complexed with tyrosine J. Mol. Biol. *194*, 287-297.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Caprara, M.G., Mohr, G., and Lambowitz, A.M. (1996a). A tyrosyl-tRNA synthetase protein induces tertiary folding of the group I intron catalytic core. J. Mol. Biol. *257*, 512-531.

Caprara, M.G., Lehnert, V., Lambowitz, A.M., and Westhof, E. (1996b). A tyrosyl-tRNA synthetase recognizes a conserved tRNA-like structural motif in the group I intron catalytic core. Cell *87*, 1135-1145.

Caprara, M.G., Myers, C.A., and Lambowitz, A.M. (2001). Interaction of the *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase (CYT-18 protein) with the group I intron P4-P6 domain. Thermodynamic analysis and the role of metal ions. J. Mol. Biol. *308*, 165-190.

Carter, P., Bedouelle, H., and Winter, G. (1986). Construction of heterodimer tyrosyl-tRNA synthetase shows tRNA^{Tyr} interacts with both subunits. Proc. Natl. Acad. Sci. USA, *83*, 1189-1192.

Cech, T.R., Zaug, A.J., and Grabowski, P.J. (1981). In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. Cell *27*, 487-496.

Cech, T.R., and Bass B.L. (1986). Biological catalysis by RNA. Annu. Rev. Biochem. *55*, 599-629.

Cech, T.R., (1990) Self-splicing of group I introns. Annu. Rev. Biochem. *59,* 543-68.

Chen, X., Mohr, G., and Lambowitz, A.M. (2004). The *Neurospora crassa* CYT-18 protein C-terminal RNA-binding domain helps stabilize interdomain tertiary interactions in group I introns. RNA *10*, 634-644.

Cherniack, A.D., Garriga, G., Kittle, J.D., Jr, Akins, R.A., and Lambowitz, A.M. (1990). Function of Neurospora mitochondrial tyrosyl-tRNA synthetase in RNA splicing requires an idiosyncratic domain not found in other synthetases. Cell *62*, 745-755.

Costa, M., and Michel, F. (1997). Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: comparison with in vivo evolution. EMBO J. *16*, 3289-3302.

Dauter Z., Dauter M., and Rajashankar K.R., (2000). Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. Acta Crystallogr. D Biol. Cryst. *56*, 232-237.

Golden, B.L., Gooding, A.R., Podell, E.R., and Cech, T.R. (1998). A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. Science *282*, 259-264.

Golden, B.L., Kim, H., and Chase, E. (2005). Crystal structure of a phage Twort group I ribozyme-product complex. Nat. Struct. Mol. Biol. *12*, 2-3.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. Bioinformatics *15*, 305-308.

Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis *18*, 2714-2723.

Guijarro J.I., Pintar, A., Prochnicka-Chalufour, A., Guez, V., Gilquin, B., Bedouelle, H., and Delepierre, M. (2002). Structure and dynamics of the anticodon arm binding domain of Bacillus stearothermophilus tyrosyl-tRNA synthetase. Structure *10*, 311-317.

Guo, Q., and Lambowitz, A.M. (1992). A tyrosyl tRNA synthetase binds specifically to the group I intron catalytic core. Genes Dev. *6*, 1357-1372.

Hall, K.B., and Fox, R.O. (1999). Directed cleavage of RNA with protein-tethered EDTA-Fe. Methods *18*, 78-84.

Haugen P., Simon D.M., Bhattacharya D. (2005). The natural history of group I introns. Trends Genet. *21*, 111-119.

Jakes R., and Fersht, A.R. (1975). Tyrosyl-tRNA synthetase from *Escherichia coli*. Stoichiometry of ligand binding and half-of-the-sites reactivity in aminoacylation. Biochemistry *14*, 3344-3350.

Kämper, U., Kück, U., Cherniack, A.D., and Lambowitz, A.M. (1992). The mitochondrial tyrosyl-tRNA synthetase of *Podospora anserina* is a bifunctional enzyme active in protein synthesis and RNA splicing. Mol. Cell. Biol. *12*, 499-511.

Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graphics *14*, 51-55.

Lambowitz, A.M., and Perlman, P.S. (1990). Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. Trends Biochem. Sci. *15*, 440-4.

Mannella, C.A., Collins, R.A., Green, M.R., and Lambowitz, A. M. (1979). Defective splicing of mitochondrial rRNA in cytochrome-deficient nuclear mutants of *Neurospora crassa*. Proc. Natl. Acad. Sci. USA *76*, 2635-2639.

Matthews, B.W. (1968). Solvent content of protein crystals. J. Mol. Biol. *33*, 491-497.

McRee, D.E. (1992). XtalView: A visual protein crystallographic software system for XII/Xview. J. Mol. Graphics *10*, 44-47.

Michel, F., and Westhof, E. (1990). Modeling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J. Mol. Biol. *216*, 585-610.

Mohr, G., Zhang, A., Gianelos, J.A., Belfort, M., and Lambowitz, A.M. (1992). The Neurospora CYT-18 protein suppresses defects in the phage T4 *td* intron by stabilizing the catalytically active structure of the intron core. Cell *69*, 483-494.

Mohr, G., Rennard, R., Cherniack, A.D., Stryker, J., and Lambowitz, A.M. (2001). Function of the Neurospora crassa mitochondrial tyrosyl-tRNA synthetase in RNA splicing. Role of the idiosyncratic N-terminal extension and different modes of interaction with different group I introns. J. Mol. Biol. *307*, 75-92.

Myers, C.A., Kuhla, B., Cusack, S., and Lambowitz, A.M., (2002). tRNA-like recognition of group I introns by tyrosyl tRNA synthetase. Proc. Natl. Acad. Sci. USA *99*, 2630-2635.

Otwinowski, Z., and Minor, M. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307-326.

Rho, S.B., Lincecum, T.L., Jr. and Martinis, S.A. (2002). An inserted region of leucyl-tRNA synthetase plays a critical role in group I intron splicing. EMBO J. *21*, 6874-6881.

Saldanha, R.J., Patel, S.S., Surendran, R., Lee, J.C., and Lambowitz, A.M. (1995). Involvement of *Neurospora* mitochondrial tyrosyl-tRNA synthetase in RNA splicing. A new method for purifying the protein and characterization of physical and enzymatic properties pertinent to splicing. Biochemistry *34*, 1275-1287.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849-861.

Winter, G., Koch, G.L., Hartley, B.S., and Barker, D.G. (1983). The amino acid sequence of the tyrosyl-tRNA synthetase from *Bacillus stearothermophilus*. Eur. J. Biochem. *132*, 383-387.

Zaug, A.J. and Cech, T.R. (1982). The intervening sequence excised from the ribosomal RNA precursor of Tetrahymena contains a 5-terminal guanosine residue not encoded by the DNA. Nucleic Acids Res. *10*, 2823-2838.

PART II: A CONTINUOUS THREE-DIMENSIONAL DNA LATTICE: TOWARD A MOLECULAR SCAFFOLD FOR CRYSTALLIZATION

Chapter 6: Introduction

6.1 OVERVIEW

Here I present the 2.1 Å crystal structure of a 13-nucleotide DNA molecule, d(GGACAGATGGGAG), that self-assembles into a three-dimensional lattice through base pairing and base stacking interactions. This DNA sequence was designed to form a duplex with a biotinylated complementary DNA strand tethered to streptavidin, as part of a tripartite complex. During crystallization trials, the oligonucleotide crystallized by itself, excluding the complementary DNA-streptavidin complex. Although the bulk of DNA oligomer crystal structures consist of short double helical oligonucleotides held together by Watson-Crick base pairing, the structure described here contains only two Watson-Crick pairs (secondary interactions), and the rest of the pairings (four base pairs) are non-Watson-Crick, falling into the category usually regarded as tertiary interactions. One feature that makes these crystals unique is the continuous base pairing between monomers. Interactions from any one monomer to any other monomer in the crystal can be established by tracing through hydrogen bonds. The presence of 4/6 non-Watson-crick base pairs is unique amongst DNA crystal structures not containing a non-Watson-Crick motif, such as G-tetrads, and is the only DNA structure to display continuous base pairing. Although this DNA crystal has not been designed rationally, it provides a unique look at alternatives to Watson-Crick base pairing in DNA self-assembly, and fortuitously solves the problem of creating a porous three-dimensional DNA lattice for potential nanotechnological applications. Finally, using this structure as a framework, I have successfully engineered three-dimensional crystals with altered lattice dimensions.

6.2 DNA IN NANOTECHNOLOGY

Rational design and assembly of nanometer size molecular objects has been a major goal of supramolecular chemistry and nanotechnology. A variety of molecular building blocks have been successfully used in the programmed assembly of nanoscale structures (*e.g*. Bong et al., 2001; Shivanyuk et al., 2001; Fenniri et al., 2001. One of the most successful of these has been DNA. Constructions of complex topology, including a cube, Borromean rings and two-dimensional tiled sheets have been created with DNA (Seeman, 2003). Several examples of nanomechanical devices constructed from DNA have also been reported (Mao et al., 1999; Yurke et al., 2000; Yan et al., 2002). One important goal of structural DNA nanotechnology has been the engineering of a three-dimensional DNA lattice (Seeman, 1982).

DNA has several features that make it an excellent building block for the construction of nanometer scale structures. First, interstrand cohesion mediated by Watson-Crick base pairing provides predictability to duplex DNA interactions and to helix geometry (Qiu et al., 1997). Base pairing provides rational programmability for DNA duplex formation in a complex structure. The DNA double helix is stiff over short distances (Hagerman, 1988), allowing for predictable lengths and orientations during assembly. DNA of designed sequence is easily synthesized through phosphoramidite chemistry (Caruthers, 1985), with both conventional nucleotides and with modified derivatives.

The Watson-Crick DNA double helix axis is topologically linear; the ability to form complex structures from DNA requires branching from these lines (Seeman, 2003). Branched DNA occurs naturally during recombination in the form of Holliday junctions and several designs have used asymmetric sequences, programmed by Watson-Crick base pairs, to create non-migratory branched junctions. Multiple-crossover motifs have been designed to self-assemble in a variety of two-dimensional arrays based on the Watson-Crick pairing of sticky ends (Winfree et al., 1993).

Biological examples of structurally complex nucleic acids are not limited to Watson-Crick interactions. Non-standard base pairs play a crucial role in the formation of the functional tertiary structures for a variety of RNA molecules (Batey et al., 1999; Westhof and Fritsch, 2000) including tRNA molecules, rRNA molecules, and RNA introns. For DNA, sequences within telomeric repeats of chromosomal DNA are known to produce non-standard base pairs, such as G-tetraplex structures (Sundquist and Klug, 1989). DNA triplet repeat expansions, implicated in a number of human diseases, have also been postulated to form non-standard pairs (Huertas and Azorin, 1996). Non-Watson-Crick base pairs are thermodynamically less favorable than canonical pairings (Peyret et al., 1999), but they appear readily in the absence of opportunities for Watson-Crick pairing or in favorable environments. Their formation, while not always predictable, provides structural diversity that in some cases is required for biological function. In the case of RNA molecules, the formation of non-Watson-Crick base pairs is key to proper folding (Westhof and Fritsch, 2000) and to RNA recognition by proteins and other ligands (Hermann and Westhof, 1999). Predictable non-Watson-Crick base pairs may provide a similar structural diversity for the creation of DNA nanoscale motifs.

6.3 APPLICATIONS OF DNA NANOTECHNOLOGY

Nanotechnology requires the ability to orient one molecule relative to another on the nanometer scale. Three-dimensional DNA crystals have a number of potential nanotechnological applications that could help to solve current problems in biology, as well as play a role in the development of molecular electronics.

Single crystal X-ray diffraction is the most powerful technique for structure determination of biological molecules, but is hampered primarily by the difficulty in producing diffraction-quality crystals. A 3-D DNA lattice could be used as a molecular scaffold for orienting biological macromolecules. Tethering guest molecules at specific sites and orientations within the molecular scaffold would integrate them into the crystal lattice, allowing their structures to be determined by X-ray diffraction techniques (Seeman, 1982; may add figure for this). In addition to improving the chances of crystallizing target molecules, this technique would eliminate much of the wastefulness associated with current crystallization protocols. By pre-assembling a crystal lattice out of DNA to serve as a scaffold, the need for screening many conditions for crystal growth is eliminated. By eliminating this “sparse matrix” screening, many target molecules could be tested in parallel, and the quantities of the target molecules required would be significantly reduced.

A DNA-based scaffold would not be limited to orienting biological molecules. Positional assembly of conductive or semi-conductive materials such as carbon nanotubes or metallic nanoparticles would open a new possibilities for developing nanoscale molecular electronics. A number biological molecules, including DNA, have been used for the assembly or as components of molecular electronics (Liu et al., 2004). Because the DNA lattice crystals described here are inherently three-dimensional, they could offer far greater positional control as well as increased density for electronic components.

Chapter 7: Crystal structure of a continuous three-dimensional DNA lattice

7.1 CRYSTALLIZATION

All oligonucleotides were synthesized using standard phosphoramidite chemistry. Commercially available phosphoramidites with bromine atoms located at the 5 position of the pyrimidine ring were used for phasing (Glen Research, California). DNA molecules were purified by polyacrylamide gel electrophoresis, electroeluted, and dialyzed against pure water. The DNA concentration was adjusted to 200 μM. Crystals were grown in batch mode by mixing 3 μL of DNA solution with 1 μL of 120 mM magnesium formate, 50 mM LiCl, and 10% 2-methyl-1,3-propanediol (MPD). Drops were incubated at 37° C overnight, after which the temperature was lowered to 27° C. Crystals shaped like hexagonal pyramids grew overnight to a size of 400 μM in the longest dimension.

7.2 DATA COLLECTION AND REFINEMENT

The lattice assembled in space group $P6_4$ with the dimensions a = b = 40.6 Å, c = 55.2 Å containing a single 13-mer in the asymmetric unit. Data were collected at National Synchrotron Light Source beamline X8-C at 100 K. Anomalous data used in SAD phasing were collected at the maximum absorption edge of Br (0.9200 Å). Data were processed in HKL2000 (Otwinowski and Minor 1997). Phasing was done by single wavelength anomalous dispersion (SAD) using the 5-bromo deoxycytidine and 5-bromo deoxyuridine modified oligonucleotide (Table 1). Refinement of heavy atom positions using anomalous dispersion was carried out in CNS (Brunger, et al., 1998). CNS was used to solvent-flatten electron density maps generated using the refined heavy atom sites or the heavy atom sites with the sign of the fractional coordinates flipped in space groups $P6_4$ and $P6_2$. The distinction of right- and left-handed helices was visible between the two space groups with clearly traceable phosphodiester backbones. The model was fit to the electron density using Xfit (McRee, 1992). The base and ribose of G13 were not present in maps generated from experimental phases or maps using combined phases from partially or fully fit models. Refinement was initially carried out in CNS to an R-factor = 25% R_{free} = 28%, at which time the refinement was switched to REFMAC5 (Murshudov et al., 1997). The same test set reflections were used with REFMAC5 to obtain the final R-factors given in Table 7.1.

7.3 STRUCTURAL FEATURES

7.3.1 Crystal packing

The overall geometry of the DNA crystal lattice can be visualized as three repeating layers of parallel helices, with each layer rotated 60° relative to the flanking layers (Figure 7.1). Each layer is composed of coaxially stacked pseudo-infinite helices separated by 20 Å of solvent space. These stacked helices interact with adjacent layers through parallel-stranded base pairing at the regions of coaxial stacking. Each DNA monomer forms base pairs with three identical strands related by crystallographic

Table 7.1. Data Collection and Model Refinement Statistics

Data set	Bromo
X-ray source	NSLS X8-C
Wavelength (Å)	0.92002
Max. resolution (Å)	2.1
Completeness (%)[a]	99.2 (99.4)
I/σ (I)	20.7 (10.7)
Multiplicity	7.0
Reflections	5516
R_{merge} (%)	7.0 (10.2)
Number of sites	2
Model refinement	
Resolution range (Å)	20-2.1
Reflections	2868
R_{work} (%)	20.3 (22.8)
R_{free} (%)	23.3 (25.8)
Geometry	
R.m.s.d bonds (Å)	0.021
R.m.s.d. angles (°)	2.8

[a] Values of the outermost bins are given in parentheses.

symmetry, yielding six unique base pairs within two regions of pairing (Figure 7.1). The helical region forms an anti-parallel double helix of B-form geometry through the base pairing of C4-G9 from two monomers. The phosphodiester backbone turns sharply between residues C4 and A3, moving G1-G2-A3 out of the axis of the helical region; this feature allows G1-G2-G3 to pair with G10-G11-A12, respectively, of a monomer in an adjacent layer. These parallel-stranded homopurine base pairs form an interlayer junction that connects each two-dimensional layer into a continuous three-dimensional array. The 20 Å spacing between the parallel helices within each layer creates an internal network of solvent channels (Figure 7.2). These channels run parallel with and perpendicular to the 6_4 symmetry axis through the length of the crystal. The channels parallel to the six-fold axis have a circular area of ~300 Å^2 in projection and a volume of ~17,500 Å^3 per unit cell. Each of the channels running perpendicular to the 6-fold axis has an area of ~360 Å^2 and a volume of ~14600 Å^3 per unit cell. These perpendicular channels intersect the parallel channel and they share ~5500 Å^3 per unit cell. The base and ribose of the 3'-terminal residue, G13, appears to be oriented into the channel intersection and could not be seen in the electron density maps.

7.3.2 Anti-parallel double helical region

The double helical region contains three unique base pairs formed by the anti-parallel base pairing of C4-G9 of two monomers (Figure 7.3). This region consists of two repeated sets of these pairings about a dyad axis between the central G6-A7 mispairs. The two outermost pairs are formed by C4-G9 and A5-T8 and are the only two Watson-Crick

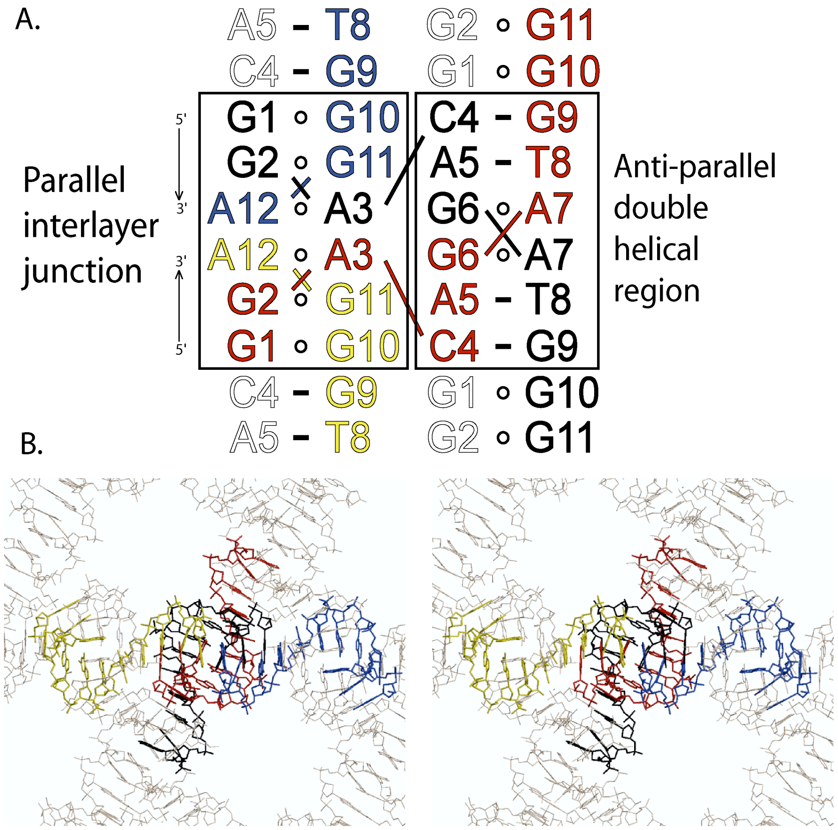

Figure 7.1. Two- and three-dimensional representations of a continuously base-paired DNA lattice. (A) The secondary structure of the 3-D DNA array highlighting the two regions of base pairing and the stacking interactions that make up the lattice. Four strands, all of which are related by crystallographic symmetry, are colored differently to highlight the base pairing relationship between them. The double helical region forms *via* anti-parallel base pairing and contains a dyad axis between the central G-A mismatch base pairs. The interlayer junction forms three parallel-stranded homopurine base pairs with two sets of these interactions stacked upon a two-fold axis. Only the phosphate of G13 was present in electron density and it has been omitted for clarity. Open circles between base pairs represent non-canonical pairings. Dashes indicate Watson-Crick base pairs. (B) A stereo image showing the three-dimensional relationship of the same four strands shown in (A) viewed down the dyad axis.

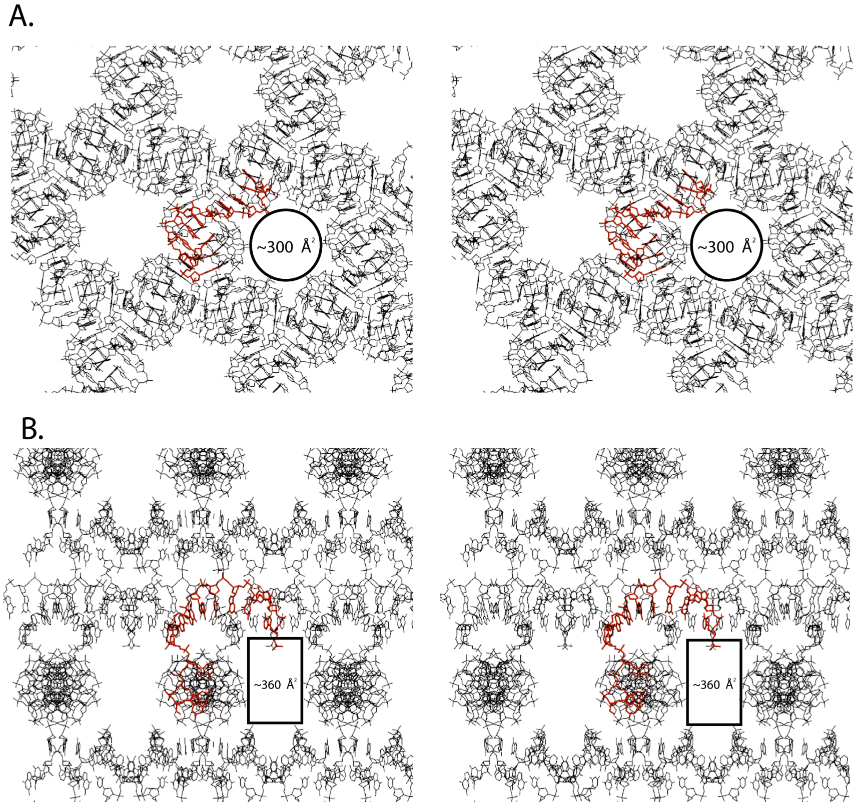

Figure 7.2. Crystal packing of d(GGACAGTGGGAG) forms solvent channels that run through the length of the crystal in four directions. (A) A stereo view down the six-fold axis showing the hexagonal channels running the length of the crystal. For reference, a monomer is colored in red. These channels have a cross-sectional area of ~ 300 $Å^2$ using the G9 phosphate-G9 phosphate distance across the channels as the diameter. (B) A stereo view orthogonal to the six-fold symmetry axis looking down one layer of coaxially stacked helices. One set of solvent channels runs parallel to each helical layer, resulting in three sets of solvent channels at 60° to each other. These channels have a cross-sectional area of ~ 360 $Å^2$. The channels in (A) and (B) intersect and share a volume of ~5800 $Å^3$. All measurements are based on atom-atom distances.

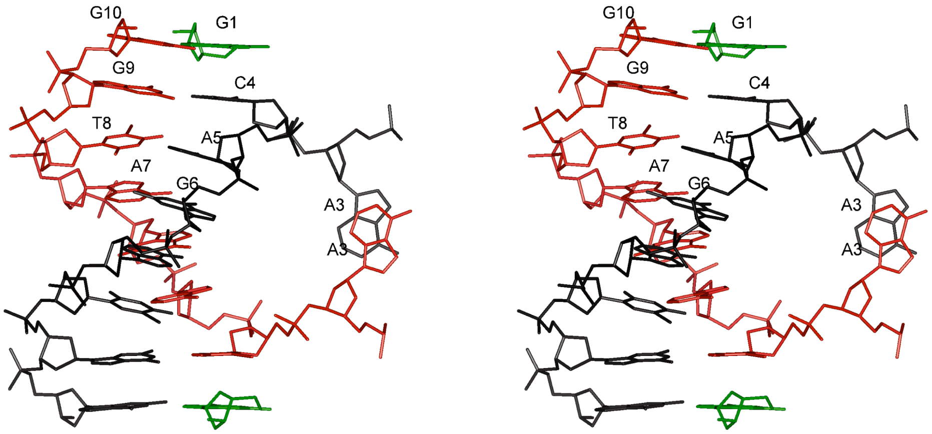

Figure 7.3. A stereo figure of the double-helical region. The helical reion contains two sets of three unique base pairs. C4 does not stack continuously with the helical region, but does stack with G1 from another monomer, shown in green. The minor groove is on the left, the major groove on the right. The major groove is partially obstructed by the sharp bend between A3 and C4.

pairs in this structure. The Watson-Crick pairs flank two sheared G6-A7 pairs. The crystal structure of the oligonucleotide d(CCAAGATTGG) contains two similar Watson-Crick-flanked sheared G-A pairs and comparison of the four central bases from the x-ray structure of this molecule (Prive et al., 1991) with the helical region reveals similar local and helical parameters. The structures conform well to canonical B-form DNA parameters and compensate for the non-Watson-Crick pair by x-displacement as determined using 3DNA (Lu and Olson, 2003). Large propeller values for the G-A pairs and the flanking A-T pairs (-25.5° and -24.0° respectively) are stabilized by a hydrogen bond between N2 of G6 and the O2 of the flanking T8. This same interstrand hydrogen bond is observed in the d(CCAAGATTGG) structure. Base stacking in the helical region is uninterrupted from A5 to G9, but a backbone distortion resulting in a large y-displacement between A3 and C4 moves the C4 base out of this helical axis, preventing intrastrand stacking with A5. This sharp bend is accompanied by a slight inclination of the C4-G9 pair, resulting in a partial closure of the major groove. While C4 does not stack with any bases from the helical region, it does stack with G1 from a molecule in an adjacent layer (Figure 7.3).

7.3.3 Interlayer region composed of non-Watson-Crick base pairs

The interlayer region is composed of two symmetry-related sets of three homopurine base pairs that stack about a two-fold axis and act as connections between helical layers, to generate the three-dimensional lattice (Figure 7.4). The outermost mismatch from this symmetry axis, G1-G10, is stabilized by two hydrogen bonds

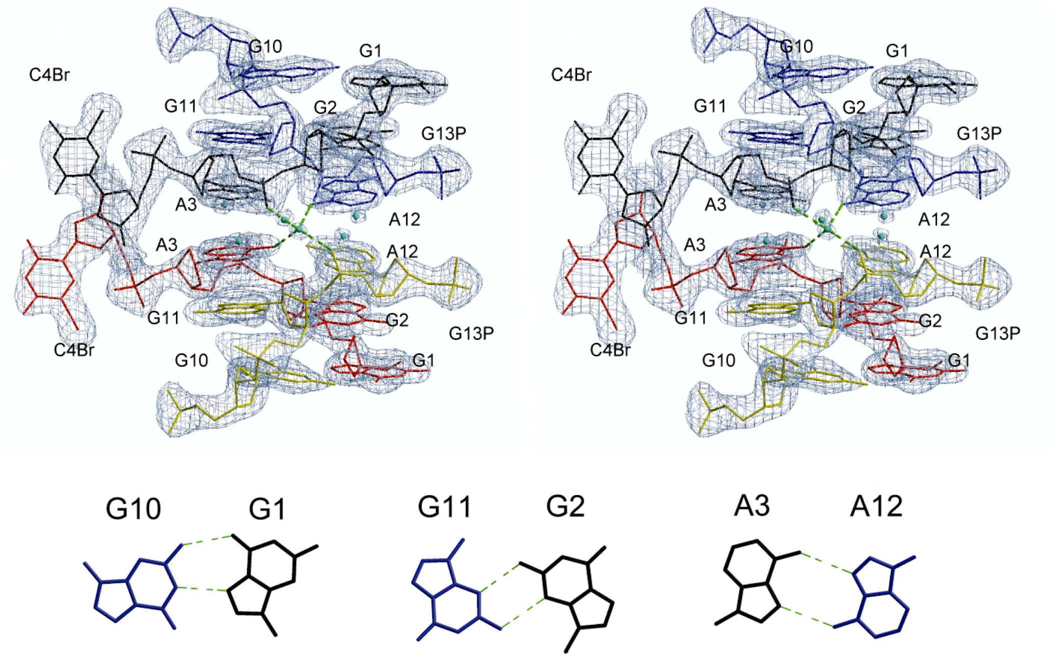

Figure 7.4. Stereo figure of the interlayer junction. Solvent flattened experimental electron density contoured at 1.5 σ. Each of the four strands making up the junction is colored differently as in Figure 7.1. Blue spheres indicate solvent molecules hydrating the junction. Dashed green lines represent the solvent-mediated hydrogen bonding between phosphate oxygen atoms with A3 and A12 bases across the two-fold symmetry axis. The lower panel images show a top-down view of the hydrogen bonding between the parallel base pairs. Figures were generated with Xfit (McRee, 1992), and Raster3D (Merritt and Bacon, 1997)

between the major groove edge of G1 and the Watson-Crick edge of G10. The intermediate pairing is formed by G2-G11 *via* symmetrical N2-N3 hydrogen bonding of the minor groove edges. The innermost pair, A3-A12, is also symmetrical, through N6-N7 hydrogen bonding. Two A3-A12 pairs stack about a two-fold axis perpendicular to the parallel-stranded helix axis formed by the ends of the four strands. This provides the end-to-end stacking responsible for generating the pseudoinfinite nature of each helical layer. Intrastrand base stacking is continuous from A5 of the double helical region through G10 of the interlayer junction, but is interrupted by the staggering of the G2-G11 pair. This base pair displays cross-strand stacking; G11 stacks with A3 from its partner strand and G2 stacks with the A12 of its partner strand, as represented schematically in Figure 7.1. Along with base stacking, solvent-mediated base-backbone hydrogen bonding provides additional stability across the two-fold axis. N6 amino groups from non-partner A3 and A12 bases form a shared hydrogen bond with a well-ordered solvent molecule. This solvent molecule serves to bridge phosphate oxygens from hydrogen bonded A3 and A12 partners. This molecule belongs to a group of well-ordered solvent molecules that hydrate the two-fold junction (Figure 7.4).

Comparison of the temperature factors between the interlayer junction and helical region suggests that the non-canonical parallel region is better ordered than the near-B-form double-helical region. The average B-factor for the six residues in the interlayer junction is 35 $Å^2$, while the average for the six residues of the double helical region is 48 $Å^2$. This ordering likely reflects the strong stacking interactions between bases in this region.

7.3.4 Comparison to other parallel-stranded DNA structures

Creation of nanoscale objects from DNA requires the predictability of hydrogen-bonding interactions between strands and the formation of a coherent local product structure from these interactions. Comparing the structure described here with parallel-stranded DNA structures containing 5'-GGA or 5'-CGA parallel regions indicates the homo-nucleotide base pairs found here can satisfy these requirements.

The three base pairs of the interlayer junction are virtually identical to the parallel-stranded homopurine pairs seen in NMR structure of d(GGAGGAT) in which two symmetrical strands make up a V-shaped ‘arrowhead’ motif (Kettani, et al., 1999) (Figure 7.5). Both molecules use identical hydrogen bond donor and acceptor pairs. The largest differences between the base pair parameters of the two structures are in the propeller and the inclination of the outermost G1-G10 pair and likely are due to stacking forces imposed by crystal formation. The 5’-terminal guanine is in the *syn* orientation allowing for Hoogsteen pairing in both structures and appears to play an important role in stabilizing the continuous array. The pyrimidine ring of G1 stacks with C4 from the helical region of an adjacent monomer and provides stacking between the parallel and anti-parallel regions of base pairing (Figures 7.1 & 7.3).

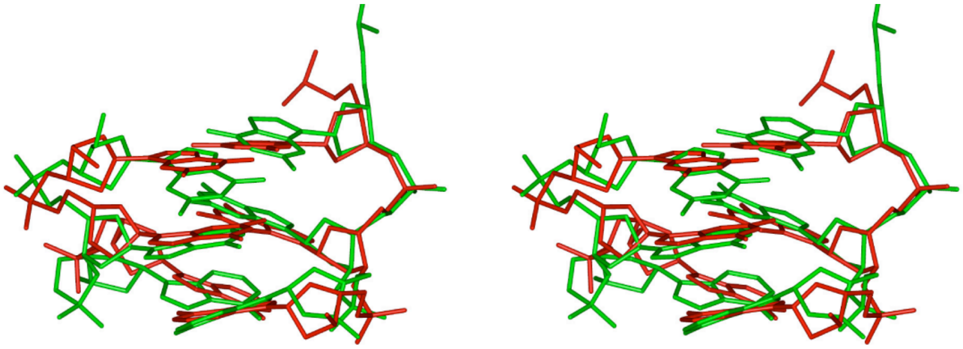

Figure 7.5. Stereo structural comparison of homo parallel base pairs. The interlayer junction is structurally similar to the V-shaped arrowhead motif. The unique pairs of the interlayer junction are in red, and a representative model of the arrowhead motif (PDB: 1B3P) is shown in green.

Several structural studies of molecules containing the DNA sequence 5'-CGA have observed similar parallel base pairing. In the case of these molecules, the C-C base pair requires that one cytosine residue be hemiprotonated to form symmetrical hydrogen bonds. NMR studies have confirmed pH dependent transitions for the molecules d(CGATCG), d(TCGATCGA), and d(CGATCGATCG) (Robinson et al., 1992). The crystal structure of d(GCGAAAGCT) displays both parallel homo base pairs and anti-parallel Watson-Crick pairs between symmetry related molecules, similar to the structure described here (Sunami et al., 2002). This structure forms a parallel region by the self-pairing of the second, third and fourth nucleotides and an additional pair from the fifth nucleotide, forming a symmetrical A-A pair *via* two N1-N6 hydrogen bonds. Superposition of the interlayer junction and 5'-CGA parallel region from d(GCGAAAGCT) show that these three base pairs are structurally equivalent, with minor differences noted in the 5' pair (Figure 7.6). The C(+)-C hydrogen bond interface is coincident with the parallel helix axis (in this case also a crystal axis), while the Hoogsteen base pairs in 5'-GGA parallel regions result in larger y-displacement values for the terminal G-G pairing. Importantly, all of these structures display interstrand stacking between the second and third base pairs and in all cases these nucleotides are G and A respectively. This strong interstrand stacking is thought to provide much of the stabilization for the parallel stranded helix (Robinson et al., 1992).

One common sequence characteristic of the arrowhead motif and the interlayer junction is the presence of a 5’-terminal GGA. Our attempts to crystallize

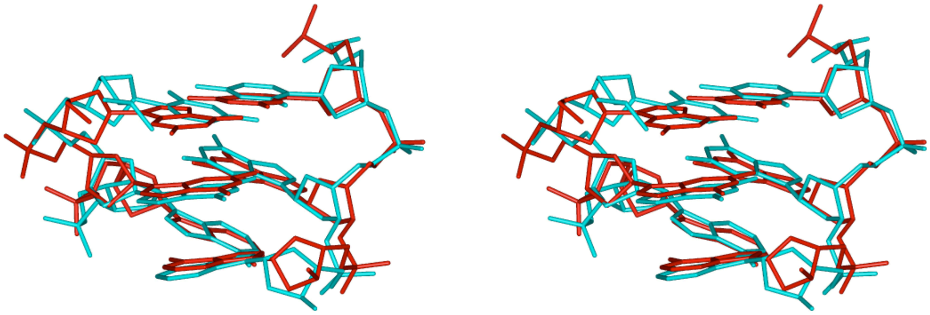

Figure 7.6. Structural comparison of 5'-CGA and 5'-GGA parallel base pairs. Homoparallel regions with the sequence 5'-GGA and 5'-CGA are structurally congruent. The three unique pairs of the interlayer junction are shown in red, and the first three parallel bases in the crystal structure of d(GCGAAAGCT) (PDB: 1IXJ) are shown in cyan. The 5'-CGA structure is symmetrical about the helical axis, while the terminal G-G Hoogsteen pair in the interlayer junction makes this region asymmetric.

oligonucleotides with an additional 5’ nucleotide were not successful. Structural studies of another DNA oligonucleotide containing internal GGA repeats, d(GTGGAATGGAAC), find an anti-parallel duplex with two unpaired guanines flanked by sheared of these first two bases may limit the ability to adopt the parallel base pairs seen in the interlayer junction and arrowhead motifs. In the case of d(GCGAAAGCT), the 5'-CGA parallel region contains one nucleotide 5’ of the CGA. The 5'-CGA structures are likely able to accommodate this extra nucleotide due to the differences in crystal packing and the additional symmetry allowed by the 5'-terminal hemiprotonated C-C pair compared with the asymmetric G-G Hoogsteen pair.

Chapter 8: Design and assembly of a scalable three-dimensional DNA lattice

To test the use of parallel homopurine base pairs in rational DNA engineering, I designed, modeled, and crystallized a two-strand DNA lattice using the base pairing interactions observed in the 13-mer structure. This lattice contains a variable length insertion between the helical region and interlayer junction that is paired with a second strand to generate one full turn of anti-parallel Watson-Crick helix (Figure 8.1).

8.1 STRUCTURE MODELING

This structure was modeled as a four-stranded dimer in the spacegroup $P3_1$ to avoid the difficulty of maintaining base pairing in the double helical region between symmetry related molecules. Building of the symmetry related molecules with a predicted unit cell that maintained all the interactions seen in the interlayer junction of the 13-mer structure showed a large expansion of the solvent channels running parallel to the six-fold axis (Figure 8.2). The circular area expands nearly 20-fold when viewed down the six-fold axis (300 $Å^2$ to 6,500 $Å^2$). This results in ~356,000 $Å^3$ of solvent space per unit cell in the channels running parallel to six-fold axis.

8.2 CRYSTALLIZATION AND INITIAL STRUCTURE CHARACTERIZATION

Crystallization trials of oligonucleotides with insertion and second strand lengths of ten and eleven nucleotides yielded hexagonal crystals within two days using the same conditions as for the single-stranded lattice (Figure 8.3). These crystals grew to a

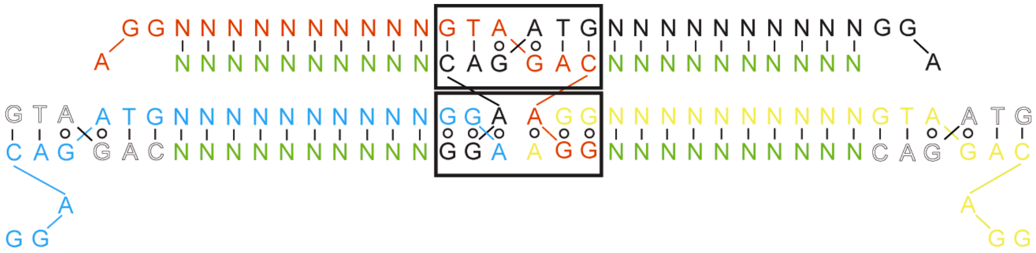

Figure 8.1. Design of an expanded three-dimensional DNA lattice. The extended DNA lattice was designed by placing a 10-11 base insertion between the helical region and the interlayer junction and using a second complementary strand (green) to extend one full helical turn. The secondary structure shows the same interactions of the helical region and interlayer junction.

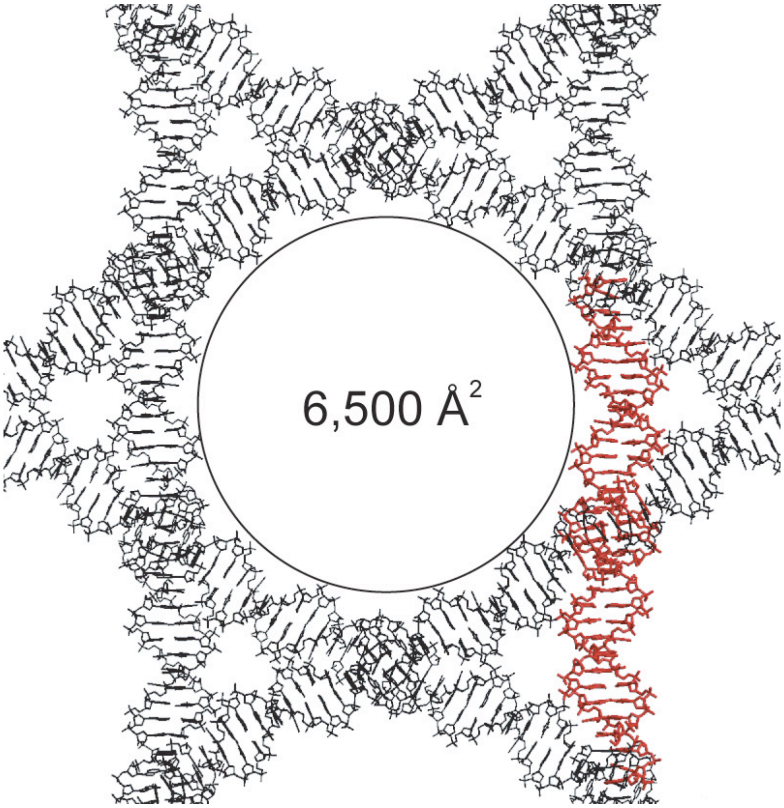

Figure 8.2. Modeling of an extended DNA lattice. Modeling of the extended lattice shows a large expansion of the solvent channels as seen down the six-fold axis. This is a nearly 20-fold expansion to 6,500 Å^2. For reference, a dimer (two extended molecules and two second strands) is shown in red.

maximum size of 0.4 mm, with the eleven base insertion crystals being consistently larger. Denaturing polyacrylamide gel electrophoresis of washed and dissolved crystals confirmed that both strands were present in the crystals (data not shown). These crystals diffracted to 5 Å resolution at room temperature (or 100 K) and exhibited high mosaicity. However, several crystals were readily indexed, confirming the primitive hexagonal lattice; similarly, they showed an experimental unit cell within 3 Å of the predicted unit cell from modeling in all dimensions. This provides excellent evidence that the model closely resembles the actual structure.

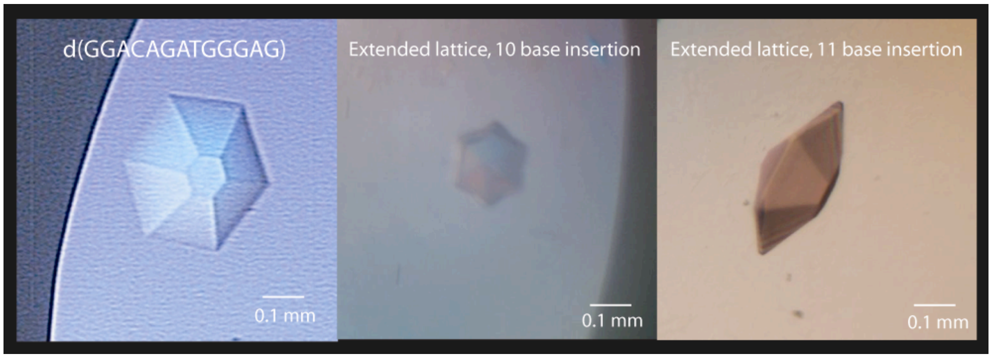

Figure 8.3. Comparison of 13-mer and expanded lattice crystals. Both crystals show hexagonal morphology and x-ray diffraction analysis confirmed the primitive hexagonal lattice and showed the experimental unit cell dimensions were within 3 Å in of the predicted unit cell as modeled in Figure 8.2.

Chapter 9: Conclusions

Continuous three-dimensional DNA crystals have been recognized as one path towards a variety of nanotechnological applications. These include uses as a scaffold for structure determination of guest molecules (Seeman, 1982), as a molecular sieve (Ribeiro, et al., 1995), and in the assembly of molecular electronic components (Robinson and Seeman, 1987). The solvent channels running the length of the DNA crystals described here make this array an attractive choice for prototyping such applications. The channels created by the 13-mer are not large enough to accommodate proteins or large macromolecules, but they are large enough to accommodate small molecules or peptides. By contrast, modeling of the extended DNA lattice indicates that the solvent channels in these crystals could accommodate globular proteins as large as 45 kDa in a single unit cell (Figure 9.1). Future work will examine if additional integer turns of Watson-Crick DNA between the double helical region and the interlayer junction can systematically increase the solvent channel volume.

The 3’ end of each 13-mer is oriented into a channel and provides a convenient point for tethering guest molecules. For the 13-mer, I have successfully produced crystals with molecules containing from one to four additional 3’-terminal nucleotides, as well as molecules containing 3’-linked fluorophores and alkylthiol groups (Figure 9.2). These crystals diffract as well as unmodified crystals, but the guest molecules could not be observed in difference density maps. This disorder is likely due to the guest molecules being present in many different orientations within the solvent channels. This finding

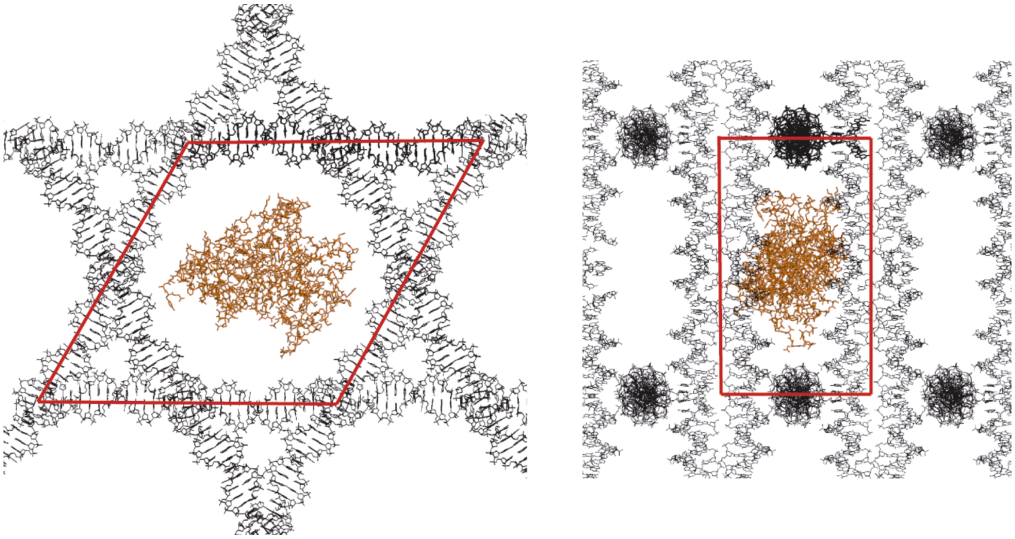

Figure 9.1. A model of the extended DNA lattice as a molecular scaffold. The large expansion of the solvent channels in the extend lattice could allow for it to be used as a macromolecular scaffold for whole proteins. Two orthogonal views show a 44 kDa fungal mitochondrial tyrosyl-tRNA synthetase fitting within a single unit cell outlined by the red boxes.

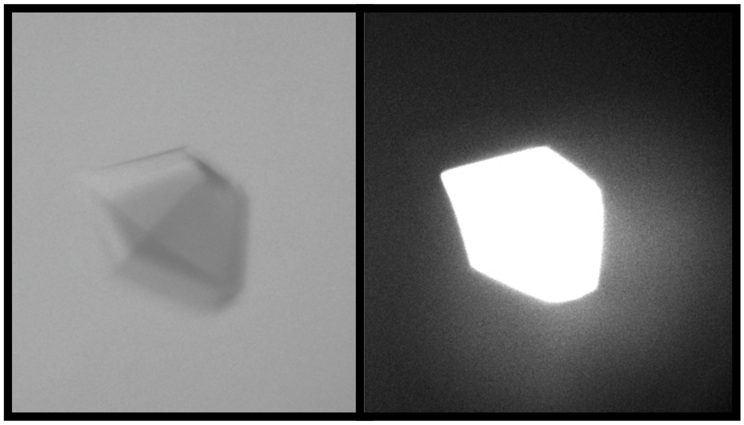

Figure 9.2. Site-specific incorporation of a fluroescent molecule into the 13-mer DNA lattice. The DNA 13-mer molecule was synthesized with a fluorescein at 3' end. The panel on the left shows a light microscope image of a crystal with the fluorescein-modified oligonucleotide. The figure on the right shows the same crystal examined with fluorescence microscopy at the excitation wavelength for fluorescein. These crystals diffract as well as unmodified crystals, but the fluorescein molecule could not be seen in difference density maps.

highlights the importance of using well-defined environments for guest molecules in a scaffolded context; in the absence of such features, disorder results. One application not dependent on constant orientation may be as a molecular sieve for filtering molecules by size or capturing them with a 3’-linked affinity tag. The programmability of Watson-Crick base pairing has been a central principle in the construction of nanoscale DNA objects, but predictable non-Watson-Crick interactions can be a potentially powerful tool for the creation of diverse DNA nanostructures (e.g., Jaeger et al., 2001). For example, the motif described here could be used as a component in nanomechanical systems, similar to the non-Watson-Crick G4 motifs used previously (Li and Tan, 2002), or as topological scaffolds (Zhu et al., 2003) with parallel features.

References

Alberti, P. and Mergny, J.-L. (2003). A DNA duplex-quadruplex exchange as the basis for a nanomolecular machine. Proc. Nat. Acad. Sci. (USA) *100*, 1569-1673.

Batey, R.T., Rambo, R.P., and Doudna, J.A. (1999). Tertiary Motifs in RNA Structure and Folding. Angew. Chem. Int. Edn. Engl *38*, 2326-2343.

Bong, D.T. and Ghadiri, M.R. (2001). Self-Assembling Cyclic Peptide Cylinders as Nuclei for Crystal Engineering. Angew. Chem. Int. Edn. Engl. *40*, 2163-2166.

Brünger, A.T. *et al*. (1998). Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. Acta Cryst. D *54*, 905-921.

Caruthers, M.H. (1985). Gene synthesis machines: DNA chemistry and its uses. Science *230*, 281-285.

Chou, S.H., Zhu, L., and Reid, B.R. (1994). The unusual structure of the human centromere (GGA)2 motif. Unpaired guanosine residues stacked between sheared G.A pairs. J. Mol. Biol. *244*, 259-268.

Fenniri, H., Mathivanan, P., Vidale, K.L., Sherman D.M., Hallenga, K., Wood, K.V., and Stowell, J.G. (2001). Helical rosette nanotubes: design, self-assembly, and characterization. J. Am. Chem. Soc. *123*, 3854-3855

Hagerman, P.J. (1988). Flexibility of DNA. Ann. Rev. Biophys. Biophys. Chem. *17*, 265-286.

Hermann, T and Westhof, E. (1999). Non-Watson-Crick base pairs in RNA-protein recognition. Chem. Biol. *6*, R335-343.

Huertas, D. and Azorin, F. (1996). Structural polymorphism of homopurine DNA sequences. d(GGA)n and d(GGGA)n repeats form intramolecular hairpins stabilized by different base-pairing interactions. Biochemsitry *35*, 13125-13135.

Jaeger, L., Westhof, E. and Leontis, N.B (2001). Tecto-RNA: Modular assembly units for the construction of RNA nano-objects. Nucl. Acids Res. *29*, 455-463.

Kettani, A., Bouaziz, S., Skripkin, E., Majumdar, A., Wang, W., Jones, R.A., and Patel, D.J. (1999). Interlocked mismatch-aligned arrowhead DNA motifs. Structure *7*, 803-815.

Li, J.J. and Tan, W. (2002). A single DNA molecule nanomotor, NanoLett. *2*, 315-318.

Liu D., Park S.H., Reif J.H., and LaBean T.H. (2004). DNA nanotubes self-assembled from triple-crossover tiles as templates for conductive nanowires. Proc. Nat. Acad. Sci. (USA) *101*, 717-722.

Lu, X-J, and Olson W.K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, Nucleic Acids Res. *31*, 5108-5121.

Mao, C., Sun, W., Shen, Z. and Seeman, N.C. (1999). A nanomechanical device based on the B-Z transition of DNA. Nature *397*, 144-146.

McRee, D.E. (1992). XtalView: A Visual Protein Crystallographic Software System for XII/Xview. J. Mol. Graphics *10*, 44-47.

Merritt, E.A., and Bacon, D.J. (1997). Raster3D: Photorealistic Molecular Graphics. Methods Enzymol. *277*, 505-524.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Cryst. D *53*, 240-255.

Otwinowski, Z and Minor, M. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods Enzymol. *276***,** 307-326.

Peyret N, Seneviratne P.A., Allawi H.T., and SantaLucia J, (1999). Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. Biochemistry *38*, 3468-3477.

Privé, G.G., Yanagi, K., and Dickerson, R.E. (1991). Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G. J. Mol. Bio. *217*, 177-199.

Qiu, H., Dewan, J.C. and Seeman, N.C. (1997). A DNA decamer with a sticky end: the crystal structure of d-CGACGATCGT. J. Mol. Biol. *267*, 881-898.

Ribeiro, F. R., Alvarez, C., Henriques, C., Lemos , F., J. M. Lopes, J.M., and M.F. Ribeiro, M.F. (1995). Structure-activity relationship in zeolites. J. Mol. Cat. A: Chem. *96*, 245-270.

Robinson, B. H. and Seeman, N.C. (1987). The design of a biochip: a self-assembling molecular-scale memory device. Protein Eng. *1*, 295-300.

Robinson, H., van der Marel, G.A., van Boom, J.H., Wang, A.H-J. (1992) Unusual DNA conformation at low pH revealed by NMR: parallel-stranded DNA duplex with homo base pairs. Biochemistry *31*, 10510-10517.

Seeman, N.C. (1982) Nucleic acid junctions and lattices. J. Theor. Biol. *99*, 237-247.

Seeman, N.C., (2003). DNA in a material world. Nature *421*, 427-433.

Shivanyuk, A. and Rebek, J. Jr., (2001). Reversible encapsulation by self-assembling resorcinarene subunits. Proc. Nat. Acad. Sci. (USA) *98*, 7662-7665.

Sunami, T., Kondo, J., Kobuna, T., Hiaro, I., Watanabe, K., Miura, K., and Takénaka, A. (2002) Crystal structure of d(GCGAAAGCT) containing a parallel-stranded duplex with homo base pairs and an anti-parallel duplex with Watson-Crick base pairs. Nucleic Acids Res. *30*, 5253-5260.

Sundquist, W.I., and Klug, A. (1989). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature *342*, 825-829.

Westhof, E. and Fritsch V. (2000). RNA folding: beyond Watson-Crick pairs. Structure *8*, R55-R65.

Winfree, E., Liu, F., Wenzler, L.A. and Seeman, N.C. (1993), Design and self-assembly of two-dimensional DNA crystals, Nature *394*, 539-544.

Yan, H., Zhang, X., Shen, Z and Seeman, N.C. (2002). A robust DNA mechanical device controlled by hybridization topology. Nature *415*, 62-65.

Yurke, B., Turberfield, A.J., Mills, A.P. Jr., Simmel, F.C. and Newmann, J.L. A DNA-fuelled molecular machine made of DNA (2000). Nature *406*, 62-65.

Zhu, L., Lukeman, P.S., Canary, J.W. and Seeman, N.C. (2003), J. Am. Chem. Soc. *125*, 10178-10179.

Vita

Paul Paukstelis was born and raised in Manhattan, Kansas. His parents, Joseph and Maria Paukstelis, were both members of the Department of Chemistry at Kansas State University. In his early years and through out his life his parents were instrumental in nurturing his interest in science and engineering. He graduated from Manhattan High School in 1993, where his interest in biology grew under the tutelage of Patrick Lamb. He attended the University of Kansas earning a B.S. in Genetics with honors in 1997. He worked in the laboratory of Dr. Dean Stetler while at the University of Kansas, beginning the transition of his interests from genetics to molecular biology and biochemistry. He spent two years as a technician in the laboratory of Dr. Donald Roufa at Kansas State University, studying RNA-mediated transcriptional regulation. He began graduate school at the University of Texas in 1999 and joined the laboratory of Dr. Jacek Nowakowski in 2000 after becoming interested in nucleic acid structure. He joined the Lambowitz lab in late 2001 where he has worked on protein-dependent group I intron splicing and initiated a collaborative effort with Dr. Nadrian Seeman of New York University to rationally design DNA-based scaffolds for structure determination.

Permanent address: 903 E. 53rd St. Austin, TX 78751

This dissertation was typed by the author.